VOL. XVIII APRIL, 1934 No. 2

The Journal of APPLIED PSYCHOLOGY

JAMES P. PORTER, EDITOR
Ohio University
WITH A BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

PHILLIP E. VERNON. The Attitude of the Subject in Personality Testing 165
RUDOLF PINTNER and GEORGE FORLANO. The Birth Month of Eminent Men 178
MILES A. TINKER. Speed in Intelligence 189
ARTHUR S. OTIS. Learning to Glide 197
ALVIN C. EURICH. Retention of Knowledge Acquired in a Course in General Psychology 209
RICHARD W. HUSBAND and JANE GODFREY. An Experimental Study of Cigarette Identification 220
VERNE W. LYON. The Use of Vocational and Personality Tests with the Deaf 224
PAUL W. TERRY. The Prognostic Value of Different Types of Tests in Courses in Educational Psychology 231
ARTHUR E. TRAXLER. Reliability, Constancy, and Validity of the Otis Intelligence Quotient 241
D. J. INGLE. A Test of Mental Instability 252
EMILY S. DEXTER and KATHARINE T. OMWAKE. The Relation Between Pitch Discrimination and Accent in Modern Languages 267
ROBERT A. DAVIS and WILLIAM A. CAMPBELL. A Correction to the Accomplishment-Quotient Technique 272
LOWELL J. SCHROEDER. The Effect of Summer Vacation on Ability in Typewriting 282
FLOYD F. CALDWELL and MARY DAVIS MOWRY. Teachers Grades as Criteria of Achievement of Bi-Lingual Children 288
T. L. ENGLE. A Personality Study of a Group of High School Honor Society Pupils 293
FRANCES V. MARKEY. Variations in Judgment 297
C. M. LOUTTIT. An Historical Note on the Application of Psychology 304
News and Notes 306
Book Reviews 307
New Books and Pamphlets 317

PUBLISHED BI-MONTHLY BY THE EDITOR
Ohio University, Athens, Ohio
Printed by
C. E. PAULEY & COMPANY, INC.
225 N. New Jersey St., Indianapolis, Indiana

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry as second-class matter at the post office, Indianapolis, Indiana.

COPYRIGHT 1933, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. Book, Indiana University, W. V. Bingham, Personnel Research Federation, New York; M. E. Broom, State Teachers College, San Diego, Calif.; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; J. E. W. Wallin, Atlantic University; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College; Henry M. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, Boston, Mass.; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; Joseph Jastrow, New School of Social Research, New York; G. M. Whipple, Danvers, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $5.00 a year. Single numbers $1.00 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

THE ATTITUDE OF THE SUBJECT IN PERSONALITY TESTING

PHILIP E. VERNON
St. John's College, Cambridge, England

INTRODUCTION

In the great majority of psychological tests, either of aptitudes or of personality traits, the subject is called on to take an active part, hence he must have some reason or incentive for so doing. If it were possible to deduce his characteristics from morphological measurements, or from chemical analysis of the blood, or even by means of the psychogalvanic reflex, there would be little need for us to concern ourselves with mental attitude to the tests. Few, however, would claim much validity nowadays for these physical or physiological methods of diagnosis. Complex psychological traits are only revealed through complex psychological techniques. Thus it is essential to obtain the subject's coöperation, and to establish a favorable subjective relationship between him and the experimental tester.

Although the testing of children raises many difficulties which are absent from adult testing, yet the technique of controlling the "subjective situation" has been quite highly developed by child psychologists such as Bronner, Healy, Cyril Burt, and many others. They lay much stress on the necessity for bringing young children into a suitable frame of mind for testing, and realize that the meaning of the test situation to the child is of greater significance than the imposition of stereotyped objective conditions. But the analogous

problems among older subjects, both in individual and in group tests, have been singularly neglected. Psychiatrists and psychoanalysts consider that the establishment of adequate contacts or "transference" between themselves and their patients is half the battle. And the importance of rapport or personal relationship in the interview (considered as a diagnostic method) is generally admitted, although definite principles and experimentally demonstrated conclusions are lacking.

In paper and pencil tests of emotional traits and attitudes, however, rapport tends to disappear. The Woodworth, Laird, Thurstone and Bernreuter inventories, and the innumerable self-rating questionnaires of introversion, or of political and religious attitudes, are standardized developments of the interview; but they differ from the interview owing to the resistance and inhibitions which tend to be evoked in most subjects who are required to divulge personal data in writing. As psychologists, accustomed to introspect, or lecture, on social, sexual, or religious sentiments and neurotic impulses, we often forget that the average college student or unsophisticated adult who undergoes these tests has been brought up to repress all mention of such topics to strangers. Even if he wishes to be sincere the subject is likely to rationalize and to deceive himself; and it is seldom that his attempts at sincerity can be checked in any way. Medical and legal psychological investigations have fully demonstrated the untrustworthiness of answers to suggestive questions about intimate personal experiences. The subject's answers are doubtless dictated, in part, not only by his notions as to what the test is meant to measure (notions which may be more or less incorrect), but also by his relations to the people who are going to see these answers. His attitude is complicated still further by his being forced to compress his natural response to each question into one of the two or three

answers which are generally provided in these standardized questionnaire tests.[1]

Free word association tests depend much less upon introspection; yet even here it has been shown that factors such as the sex of the experimenter[2] or the presence of a group[3] affect the subject's responses. Very complex attitudes are likely to be aroused also by behavior tests of temperament and character, such as the Will-Temperament tests, Moore and Gilliland's "shifty eye" test of aggressiveness, or Fernald's test of endurance and its numerous progeny. The subject must have some motive for undertaking such apparently pointless activities as writing "United States of America" at minimum speed, or keeping his eyes fixed on those of the experimenter, or standing on his toes. His behavior will depend on his personal relations with the experimenter and his speculations about the object of the test. Usually this object is concealed from him, for otherwise he would, of course, respond in accordance with his ideas of the desirability of the trait. The experimenter may have to resort to tricks, as in most tests of moral qualities, or in Morgan and Hull's test of persistence, where the subject is given a maze to trace which, unknown to him, cannot be solved. Should he discover the trick his normal reactions would be utterly distorted.

[1] Another significant factor has been pointed out by Vetter in connection with his test of radical-conservative attitudes. He finds that the degree of radicalism of the average subject who takes the test in a group is different from that of the average subject who answers the same test in solitude. Cf. Vetter, G. B., "The Measurement of Social and Political Attitudes and the Related Personality Factors." *J. Abn. & Soc. Psychol.*, 1930, 25 149-189.

[2] Cf. Cason, H. & Cason, E. B., "Association Tendencies and Learning Ability." *J. Exper. Psychol.*, 1925, 8, 167-189.

[3] Cf. Allport, F. H., "The Influence of the Group upon Association and Thought." *J. Exper. Psychol.*, 1920, 3, 159-182.

Hartshorne and May have recognized the importance of this topic in the criteria which they present for a good test of character traits.[4] These are some of their conditions. "The test situation should be as far as possible a natural situation. It should also be a controlled situation." "There should be nothing about it (the test) to trick an honest subject into an act he would repudiate if he were aware of its import." "The test should not put the subject and the examiner in false social relations to one another. The examiner should guard against being deceptive himself in order to test the subject." "The activity demanded of the subject in taking the test should have real values for him whether he is aware of these values or not."

The methods of establishing these "real values" constitute the chief problem of the present paper. We will consider the various incentives or factors which have been employed for producing favorable coöperation on the part of subjects, particularly on the part of psychologically unsophisticated college students and adults.[5] Although these incentives are invariably mixed in actual practice, they may be abstracted and broadly classified as follows:

1. Competition
2. Economic reasons
3. Miscellaneous obligations
4. Personal interest
5. Interesting material
6. Indirect incentives

[4] Hartshorne, H. and May, M. A., *Studies in Deceit.* New York, 1928.

[5] The motivation of children as subjects will be omitted here, partly because more has already been written on the topic, and partly because the present writer cannot claim personal experience in child testing. Psychologically sophisticated graduate students, who mutually act as subjects in one another's experiments, will also be left out of account; for they are generally rather unsuitable for researches in personality. The psychoanalyst has realized that people who claim to be acquainted with, and sympathetic toward, psychoanalysis actually offer the greatest "resistance" to psychoanalytic treatment. Probably the same is true of personality testing; and even if the analogy is unfair, their motivation would be different from that of unsophisticated adults.

I. Competition

In many motor tests (tapping, pursuit meter, etc.), and in tests of abilities, competition between different subjects is quite an effective incentive. But in tests which bear more directly upon personality, rivalry is usually discouraged. Indeed it is difficult to think of any valuable method of diagnosis which makes use of maximal performances and achievements. Free word association experiments demand that the subject shall produce the first word that comes into his mind. But he should not be allowed to gain the notion that the quicker he answers, the better; for then he would be likely to reply to the stimulus words immediately with non-revelatory responses which would not be true associations. Downey's Freedom from Load, Coordination of Impulses, and other Will-Temperament tests are based on maximal writing speed under various conditions. There is a distinctive competitive element in the attempt to write several words on a very short line, to copy as well as possible, to write as slowly as possible, and so forth. Many of the Bernstein and Lankes perseveration tests involve speeded and accurate performances, *e. g.*, mirror drawing, writing capital letters backward, etc. But such tests can seldom arouse much "urge to mastery;" they are not sufficiently interesting, and in general the subjects are not told their scores. Moreover, it has yet to be proved that these *are* tests of personality or temperament. At present it seems doubtful whether they possess much predictive value for the qualities which they are supposed to measure.

II. Economic Reasons

Though the tester does not bribe college freshmen with candy (as is sometimes necessary in studying the personalities of young children), many economic incentives will come to mind. Some experimenters, including

the present writer, have paid their subjects at a standard rate. Many colleges give academic credits to students who spend so many hours in acting as subjects. Sometimes the subjects can pick up hints on technique, etc., which may be useful to them in passing their own examinations. Actually the monetary incentive seems to have little but negative value. If it is absent the subject may regret the waste of time in the laboratory. If it is present he is at least contented to stay in the laboratory during the appointed hour, since the tests are not likely to constitute any more disagreeable a means of earning money than, say, working as a waiter in a restaurant. But this is no guarantee of his conscientious coöperation throughout the session. And clearly this incentive should be discouraged owing to its very different effects on persons whose incomes differ. If the subjects are drawn only from among students who greatly require money, an unfortunate selective factor (which is very likely correlated with the test variables) may operate.

Similarly it is quite impossible to control either the strength of the incentive or the representativeness of the sampling if the tester relies on members of a class in elementary psychology who are given credits for acting as subjects, or who think that they may win the professor's approval by volunteering.

III. Miscellaneous Obligations

Student subjects may often be enlisted who do not expect pecuniary or other gain and the investigator may draw upon his personal friends and acquaintances. Here an important incentive to coöperation is what Dodge calls, "the unanalyzed mass of obligations,"[6] including the subject's respect for, or friendship with, the investigator. It is a very undependable one, since the mental

[6]Dodge, R., "Conditions and Consequences of Human Variability." *Yale University Press*, 1931. Pp. 24-25.

processes in the minds of different subjects are likely to be so complex and diverse. Each subject has given his word that he will take the tests; but how far will this promise go in motivating him to answer sincerely a question about his sexual life, or, as in Landis' experiments, to cutting off the head of a live rat?

In "deep" psychiatric analyses we know that transference is essential. But in experimental investigations of personality the best relationship would seem to be ordinary friendly acquaintanceship, which is not wholly impersonal. It is worth the investigator's while to encourage casual conversations at the beginning and end of experimental sessions. For instance, during a prolonged research with college students, he may follow their athletic careers and discuss extra-curricular matters, without aiming at too great an intimacy. Undoubtedly the better he knows the subjects, the better will he be able to understand and interpret the results and the apparent inconsistencies arising out of the tests. On the other hand, it is dangerous to be too friendly with them. If he is very familiar and informal, they will often fail to take the tests seriously, and will frequently want to stop and talk about their performances. Moreover, when experimenter and subject possess many close mutual acquaintances, inhibitions are likely to be aroused through the subject's irrational fear (possibly unformulated) that the experimenter may compromise him. If, for example, the experimenter is a friend of the subject's family, the subject may be very unwilling to reveal psycho-neurotic symptoms and the like, however much he is assured that his answers are confidential.

During the actual session it is difficult to decide whether the experimenter should assume a persuasive, submissive attitude, or whether he should try to dominate the subjects into following his directions. In individual testing he will probably vary his methods to suit the particular subject, in spite of the departure from

objectivity thereby involved. But in group testing all the subjects must receive the same instructions, and most testers preface the application of pencil and paper tests with some sort of request for coöperation, whose wording must be carefully chosen to appeal to as many members of the group as possible. A professor who asks his class to answer a certain test for him can hardly expect good results if the class either despises him for his inefficiency or bashfulness, or fears him for his domineeringness and austerity.

IV. Personal Interest

Undoubtedly the most powerful and at the same time the most hazardous incentive is that of self-interest in the results. During the past five years the present writer has applied a variety of individual tests of personality to about 120 persons between the ages of 17 and 25; and he noticed that almost all of them were anxious to find out anything they could about themselves. Indeed the more a subject thinks he will be told, the more readily he does the test. Hence the writer seriously advises that, in any lengthy investigation, a number of tests should be included which may be of little value to the experimenter's own scientific aims, but which will be of interest and practical value to the subjects. Vocational aptitude tests are worth applying if they do nothing more than afford the experimenter something he can tell to the subjects. Unpleasant questionnaires on emotional traits and the like can, as it were, be smuggled in and will be answered all the more sincerely if the subject is allowed to forget about them in his interest in other aspects of his personality. (This is an example of an indirect incentive, *cf.* Section VI.)

The danger of this incentive lies in the fact that the experimenter is forced to assume the rôle of psychiatric and vocational adviser, whether or not he has had any training in these fields. He cannot avoid this difficulty

simply by posting the results of all the subjects on a notice board, where anyone can see, or by reading them out in class. Candid responses to the tests will certainly be inhibited if the subject fears that he may be announced as the most radical, the least artistic, or the most unsociable in the group. A personal interview with each subject is essential. And it is not enough simply to tell him his standing on each test, since he is almost certain to misinterpret the significance of his scores. The chief part of such an interview should consist in dissuading him from attaching too much importance to unreliable tests which may be very poor measures of the trait-names by which they are called. For these reasons the experimenter should be not only a psychometrist but also a sympathetic and experienced clinician. If he desires to investigate the social, sexual and other emotional characteristics of his subjects, it is incumbent on him not to trifle with his material. He may do incalculable harm to a subject by treating him as a mere numerical aggregate of emotional symptoms.

V. Interesting Material

Although some subjects will regard any experimental situation or test material as trivial or treat it facetiously, the majority do not seem to find a considerable intrinsic interest in most personality tests. Apart from any other incentive many persons like to introspect, within limits. Of course there are large individual differences which ought to be taken into account in evaluating the results of an introspective test. In most paper and pencil tests of attitudes the questions that are asked on economic, moral or religious opinions will generally be answered carefully without the need of any additional stimulus. Several tests which throw light both on aptitudes and on temperament, such as the Rorschach inkblots, Morgan and Hull's maze, formboards or picture completion, and McAdory's Art Judgment test, certainly attract most

subjects. On the other hand the test material may easily become boring; even the McAdory and Meier-Seashore art tests seem too long to give at one sitting. Most of the more elementary tests such as reaction time, dotting machine, perseveration, etc., employ such tedious material that personal obligations or economic incentives usually have to be called in, with consequent lack of control of results. The 125 questions of the Bernreuter Inventory, or the 223 questions of the Thurstone Schedule, even the 100 free association words of the Kent-Rosanoff list, are likely to pall long before the end. We are entirely ignorant as to the effects of excessive length, although it should be quite feasible to determine experimentally whether better results are obtained when, say, 60 minutes, 30 minutes, or 10 minutes are spent on one kind of a test at a time.[7]

VI. Indirect Incentives

However competently the experimenter may apply all or any of the above incentives, there remains still a fundamental difficulty which cannot be evaded, namely—the subject who knows that his personality is under investigation cannot react normally. Graphologists have realized this point better than have mental testers; for they say that handwriting specimens are of no use for personality diagnosis unless they have been written

[7] A related problem which requires investigation is the effect of format, spacing of print and position of answers on the results of paper and pencil tests of personality. In multiple choice or paired comparison tests of abilities this effect is known to be considerable. One step has been taken by constructors of graphic rating scales, who generally place the desirable end of successive traits sometimes on the right hand, sometimes on the left hand side of the rating sheets. Katz and Allport, in their work on *Students' Attitudes* (1930) managed to avoid space errors in their questionnaire by dividing their group of about 4000 subjects into two halves, and reversing for one-half the order of the answers to each question.

under natural and spontaneous conditions. The more the writer is conscious of the act of writing, of the shape of his letter-formations, and of his style, the less relation does the resulting script bear to his true personality. Similarly it should be recognized that every direct approach to personality, by making the subject self-conscious, introduces artificiality into the test situation. Hence it is desirable to make the most of these approaches which do not require any direct motivation at all. The observational techniques of D. S. Thomas, W. C. Olson, F. L. Goodenough and their coworkers are far the most promising of the methods which fall under this heading. Unfortunately they have found very little application so far among adult subjects. Owing to distorting subjective factors, rating methods are a poor second best. Both of these depend upon records or judgments made by external observers of the subject's unselfconscious behavior under conditions of real life.

Another indirect approach is to set the subject written or performance tests of mental or manipulative abilities which merely require the incentive of competition to yield consistent and reliable results. Certain data of significance for personality may then be deduced, quite unknown to him, either through observations of his behavior or through measurements of the variability and other incidental aspects of his performances. The free association test, when treated by Wyman's technique, gives indirect measures of interests which may someday turn out to be quite useful, though their validity is very low at present. In Strong's blanks, also, the scores for occupational interests are arrived at without direct questioning of the subject. In G. B. Watson's test of Fair-mindedness, the subject records his opinions on various controversial topics, and does not realize that he is at the same time showing up his prejudices. Here the indirect incentive is No. V, *i. e.*, interesting material. The disadvantage of some of these indirect approaches is

that they approximate to the "trick" tests which were condemned earlier in this paper. Both the Strong and the Watson tests would give false results if the subjects found out what they are supposed to measure. Nevertheless it should be possible to develop several other techniques along these lines without "putting the subject and the examiner in false social relations to one another."

CONCLUSION

This discussion of incentives is, admittedly, speculative, owing to the extreme paucity of facts. (In two recent summaries of the literature on motivation, only two references could be discovered which had any bearing on the topic.)[8] Many testers may object that a test should be judged, not by this *a priori* and subjective reasoning, but by its quantitative and objective results. If the pragmatic, empirical approach is strictly followed out, the tester has to forget that he is dealing with individual human beings; he must treat his subjects as S-R mechanisms to each of whom precisely the same test material is applied under precisely the same objective conditions. The reliability and validity of the test is then determined by the usual statistical methods. Thus the intermediate stage, namely the meaning of the test for the subject and the psychological factors which motivate his response, is left out of account.

There are two counter-arguments to this objection. First, the mentality of the subject is invariably considered by the psychologist who devises the test. He has to employ subjective mental processes, introspection and anthropomorphic interpretation, in constructing the test situations which will, he hopes, reveal the trait that he desires to measure. Moreover, as we have already seen,

[8] Cf. Hurlock, E. B., "The Psychology of Incentives." *J. Soc. Psychol.*, 1931, 2, 261-290. Diserens, C. M. and Vaughn, J., "The Experimental Psychology of Motivation." *Psychol. Bull.*, 1931, 28, 15-65.

experimenters who apply individual tests always do vary the objective conditions to a certain extent, in order to achieve more favorable subjective conditions in their testees. Secondly, this objection recoils on the heads of its proponents, since it must be acknowledged that, in spite of years of patient research by hundreds of investigators, the predictive validity of almost all personality tests so far devised is disappointingly low. It is possible that better results might be obtained if more attention were paid to what the subject thinks about the test.

THE BIRTH MONTH OF EMINENT MEN

RUDOLF PINTNER and GEORGE FORLANO
Teachers College, Columbia University

In two previous articles[1] we have reported results of the distribution of the IQ's of children according to their month of birth. In the second study a lower IQ for the winter months (January, February, March) appeared practically in all samplings, although the difference of 1.7 in IQ points was not great. The present study was undertaken to answer the question as to whether a similar phenomenon would show itself in the birth month of eminent men. If children born in the winter months are slightly lower in intelligence than those born during the rest of the year, then we might find somewhat fewer eminent men born in winter than in other seasons. Previous studies in this field which we have seen have not led to any definite conclusions. The number of cases dealt with has generally been rather small. It seemed to us desirable to study the problem further with different and larger samplings.

Sources of Data

Our samples were taken from *Who's Who in America,* Volume 17, 1932-33; *American Men of Science,* Fourth Edition, 1927; Plarr's *Men and Women of the Time,* Fifteenth Edition, London, England, 1899; and Cox's *Genetic Studies of Genius,* Volume II. 1926. The number of cases in each sampling is as follows:

[1] Pintner, R., "Intelligence and Month of Birth, *J. Appl. Psychol.,* 1931, 15, 149-154; Pintner, R., and Forlano, G., "The Influence of Month of Birth on Intelligence Quotients," *J. Educ. Psychol.,* 1933, 24, 561-584.

Who's Who in America. First sample	5,968
Who's Who in America. Second sample	6,306
American Men of Science. First sample	4,409
American Men of Science. Second sample	4,508
American Men of Science. Starred cases	1,374
Plarr's Contemporary European	2,330
Historical Geniuses—Cox's cases	271
Total	25,166

The sampling procedure consisted of tabulating the birth months of all names on certain pages throughout the volumes, for example, pages 240 to 339; 740 to 839; 1240 to 1339; and so on. Two samples in the two major sources were taken in order to check on any chance trend that might appear in any one sample. The starred cases in *American Men of Science* were tabulated separately in order to see whether the more eminent men, as judged by their contemporaries, would show an intensification of any trend that might appear among the less eminent. The geniuses studied by Cox were tabulated according to the estimated IQ as given in her work, in order to see whether such a tabulation would bear any resemblance to our findings in a previous study of the IQ's of children. The number of cases is, however, very small when distributed over twelve months, or even when grouped into four seasons.

The Results

Tables 1 to 9 give the distributions by birth month for various samplings and combinations of samplings. Table 1 shows the two samples of men taken from *Who's Who in America*. The number of cases for each month is given and also the percentages for each month. Because the months do not all contain an equal number of days, the number of cases per month has in each case been divided by the number of days in the month and this figure we have called the daily rate. This daily rate is not to be preferred to the percentage, but it allows a

TABLE 1

Who's Who in America

Two Samplings

Month of Birth	Number of Cases		Percentages		Daily Rate	
	First Sample	Second Sample	First Sample	Second Sample	First Sample	Second Sample
January	469	545	7.86	8.64	15.1	17.6
February	430	535	7.21	8.48	15.2	18.9
March	490	531	8.21	8.42	15.8	17.1
April	518	487	8.61	7.72	17.2	16.2
May	468	500	7.84	7.93	15.1	16.2
June	463	459	7.76	7.28	15.4	15.3
July	585	506	9.80	8.02	18.9	16.3
August	462	558	7.74	8.85	14.9	18.0
September	587	524	9.83	8.31	19.7	17.4
October	537	585	9.00	9.28	17.4	18.8
November	489	566	8.19	8.97	16.3	18.8
December	470	510	7.87	8.09	15.2	16.4
Total	5968	6306	99.92	99.99	16.3	17.2

more direct comparison of month with month, or season with season, having made allowance for differences in the number of days.

The two samples in Table 1 are very similar. The combination of the two samples in Table 2 shows the largest percentage of eminent men born in October, the largest daily rate in September. With reference to the seasons, summer (July, August, September) shows the largest percentage. However, the striking thing about Tables 1 and 2 is not the differences but the similarities in percentages and rates from month to month and from season to season. The largest difference between any two months is only 1.6 per cent, and the largest difference between any two seasons is only 2.7 per cent. Obviously no month or season can be said to be contributing a preponderance of eminent men.

Tables 3 and 4 show the results for the two samplings from *American Men of Science*. A study of these tables

TABLE 2

Who's Who in America

Total Cases

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	1014	8.29	32.8			
February	965	7.88	34.1	3000	24.5	33.2
March	1021	8.35	33.0			
April	1005	8.21	33.4			
May	968	7.90	31.2	2895	23.6	31.8
June	922	7.53	30.8			
July	1091	8.92	35.2			
August	1020	8.34	33.0	3222	26.3	35.0
September	1111	9.07	37.1			
October	1122	9.18	36.2			
November	1055	8.62	35.1	3157	25.7	34.2
December	980	8.00	31.6			
Total	12274	100.29	33.5	12274	100.1	33.5

shows also the similarities of the two samples and the similarities in percentages and rates from month to month and from season to season. The largest differences in percentages are still less than in the previous tables. In Table 4 the largest percentage difference between any two months (September and December) is only 1.1, and the largest difference between any two seasons (Summer and Spring) is only 1.5 per cent.

Table 5 is a combination of Tables 2 and 4, showing the combined samplings from *Who's Who in America* and *American Men of Science.* There are no marked differences between any two months or any of the four seasons.

Table 6 shows the tabulation of birth months for those individuals in *American Men of Science* who are marked with a star, denoting special eminence in their respective fields. This tabulation was made to see if the more

TABLE 3

American Men of Science

Two Samplings

Month of Birth	Number of Cases		Percentages		Daily Rate	
	First Sample	Second Sample	First Sample	Second Sample	First Sample	Second Sample
January	389	354	8.82	7.85	12.5	11.4
February	348	396	7.89	8.77	12.3	14.0
March	388	376	8.80	8.34	12.5	12.1
April	335	384	7.60	8.50	11.2	12.8
May	369	350	8.36	7.76	11.9	11.3
June	371	362	8.41	8.02	12.4	12.0
July	377	367	8.55	8.14	12.2	11.8
August	372	399	8.44	8.85	12.0	12.8
September	372	410	8.44	9.10	12.4	13.6
October	397	381	9.02	8.45	12.8	12.3
November	353	373	8.01	8.27	11.7	12.4
December	338	356	7.66	7.89	10.9	11.5
Total	4409	4508	100.30	99.94	12.1	12.3

eminent men would emphasize any trend that might be observed among the less eminent. The percentages from month to month fluctuate more markedly because of the small number of cases. The largest percentage difference between any two months (March and June) is 3.3, and between any two seasons (Winter and Spring) is 5.6. Here the differences are larger, but the trend is not the same as in the other tables, except in so far as Spring shows the lowest percentages. The high percentage of cases born in Winter is not found in any of the other distributions. It is probably a chance occurrence.

Table 7 is a combination of all cases discussed so far, and shows no definite superiority for any month or season.

Table 8 shows the distribution of eminent European individuals. It is strikingly like the tables for eminent Americans. The differences in percentage between months and seasons are all insignificant.

TABLE 4

American Men of Science

Total Cases

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	743	8.32	24.0			
February	744	8.34	26.3	2251	25.2	24.9
March	764	8.56	24.6			
April	719	8.06	24.0			
May	719	8.06	23.2	2171	24.3	23.9
June	733	8.21	24.4			
July	744	8.34	24.0			
August	771	8.65	24.9	2297	25.8	24.9
September	782	8.78	26.1			
October	778	8.72	25.1			
November	726	8.14	24.2	2198	24.6	23.9
December	694	7.77	22.4			
Total	8917	99.95	24.4	8917	99.9	24.4

Finally, Table 9 shows the distribution of all the cases, American and European, totalling 24,895. The percentage difference between any two months is never greater than 1.24, and this is certainly too small a difference to denote any marked superiority for any month. The largest percentage of eminent men is found in Summer, but Winter and Autumn have very similar percentages. Spring has the lowest percentage, being 2 per cent below the high percentage of Summer, but this difference is negligible.

Estimated IQ's of Geniuses

Cox's study of very famous historical characters gives an estimated IQ for 271 cases. We have made a distribution of these by month of birth and have calculated the mean IQ and standard deviation for each month, as well as the percentage of cases per month. These results are shown in Table 10. The highest average IQ falls

TABLE 5

Who's Who in America and American Men of Science

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	1757	8.30	56.6			
February	1709	8.06	60.4	5251	24.8	58.2
March	1785	8.43	57.6			
April	1724	8.15	57.5			
May	1687	7.97	54.5	5066	24.0	55.7
June	1655	7.82	55.2			
July	1835	8.67	59.2			
August	1791	8.48	57.9	5519	26.1	59.9
September	1893	8.95	63.1			
October	1900	8.99	61.3			
November	1781	8.42	59.4	5355	25.3	58.2
December	1674	7.92	54.0			
Total	21191	100.16	57.9	21191	100.2	57.9

TABLE 6

American Men of Science

Starred Cases Only

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	129	9.40	4.2			
February	121	8.81	4.2	388	28.2	4.31
March	138	10.03	4.4			
April	110	8.00	3.7			
May	108	7.86	3.5	311	22.6	3.4
June	93	6.77	3.1			
July	107	7.79	3.5			
August	108	7.86	3.5	330	24.0	3.6
September	115	8.36	3.8			
October	124	9.03	4.0			
November	101	7.35	3.4	345	25.1	3.8
December	120	8.73	3.9			
Total	1374	99.99	3.9	1374	99.9	3.9

in June and the lowest in July. The difference is 11.4 points in IQ, but it is only 1.72 times the S. D. of the difference. The largest average IQ for the four seasons falls in spring and the lowest in summer. This difference is only 1.52 IQ points, and it is only .47 times the S. D. of the difference. Because of the small number of cases at our disposal we can place no confidence in the reliability of our results. The percentages show a high point in January and a low point in June. A slightly larger percentage of cases falls in January than in any of the other seasons.

This analysis of the estimated IQ's of 271 geniuses has thus led to negative findings. We find no support here for our previous finding of a relatively lower IQ among children born in the winter.

CONCLUSIONS

Our data do not show any reliable differences between months or seasons with reference to the birthdays of eminent men. There is little consistency in the trend from sample to sample. The highest percentage may fall in almost any season or any month. What little consistency there is would seem to be with reference to spring, where we find the lowest percentages in almost every sampling. This trend, which is very unreliable, is directly contrary to Blonsky's[2] finding of the highest average IQ for children born in spring. It also is not in conformity with our own previous findings of the lowest average IQ for children born in the winter months.

It was this previous finding of ours, namely a slightly lower IQ for children born in winter, which led to the present study. We wondered whether the difference in IQ between winter and the other seasons would be reflected in later life by showing a smaller percentage of

[2] Blonsky, P. P., "Früh- und Spätjahrkinder,"*Jahrbuch f. kinderheilkunde*, 1929, 124, 115-120.

TABLE 7

Who's Who in America and American Men of Science

Including Starred Cases

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	1886	8.37	60.8			
February	1830	8.13	64.6	5639	24.9	62.3
March	1923	8.55	62.1			
April	1834	8.15	61.2			
May	1795	7.98	57.9	5377	23.8	59.0
June	1748	7.76	59.5			
July	1942	8.63	62.6			
August	1899	8.43	61.2	5849	25.9	63.3
September	2008	8.93	66.8			
October	2024	9.00	65.3			
November	1882	8.36	62.4	5700	25.3	61.9
December	1794	7.98	57.8			
Total	22565	100.27	61.6	22565	99.9	61.6

TABLE 8

Eminent European Contemporary Individuals

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	201	8.62	6.5			
February	183	7.85	6.5	591	25.3	6.5
March	207	8.87	6.7			
April	179	7.68	6.0			
May	192	8.23	6.2	546	23.4	6.0
June	175	7.51	5.8			
July	214	9.18	6.9			
August	192	8.23	6.2	581	25.0	6.3
September	175	7.51	5.8			
October	209	8.97	6.7			
November	209	8.97	7.0	612	26.3	6.6
December	194	8.32	6.3			
Total	2330	100.04	6.4	2330	100.0	6.4

TABLE 9

Eminent Men

All Sources Combined

Month of Birth	Number	Percentage	Daily Rate	Seasons		
				Number	Percentage	Daily Rate
January	2087	8.37	67.5			
February	2013	8.07	71.2	6230	25.0	69.0
March	2130	8.55	68.7			
April	2013	8.07	67.1			
May	1987	7.98	64.1	5923	23.8	65.1
June	1923	7.72	64.1			
July	2156	8.65	69.4			
August	2091	8.38	67.4	6430	25.8	69.8
September	2183	8.77	72.9			
October	2233	8.96	71.9			
November	2091	8.38	69.7	6312	25.3	68.6
December	1988	7.99	64.1			
Total	24895	99.89	68.2	24895	99.9	68.2

eminent men born in winter. The assumption here is that a high IQ in childhood is closely associated with eminence in later life and in particular with the type of eminence which leads to inclusion in such books as *Who's Who in America* and *American Men of Science*. Furthermore, we must remember that with children the difference between winter and other seasons was not more than two points in IQ, and this small difference may be washed out in later life so far as eminence is concerned because of the numerous other factors, in addition to general intelligence which play a part in leading to inclusion in our books of contemporary eminent men.

TABLE 10

Estimated IQ's of Geniuses by Month of Birth

Month of Birth	Mean IQ	S.D.	n	Per Cent
January	138.9	13.4	31	11.41
February	137.7	15.3	26	9.58
March	137.1	17.4	19	7.02
April	134.6	19.2	24	8.86
May	139.8	17.1	27	9.96
June	142.3	22.4	15	5.54
July	130.9	13.3	17	6.27
August	140.4	18.8	26	9.58
September	136.3	16.6	23	8.48
October	138.0	22.6	20	7.38
November	140.0	19.5	22	8.11
December	136.9	11.2	21	7.75
			271	99.94
Seasons				
Winter	138.0	15.1	76	28.0
Spring	138.5	19.4	66	24.4
Summer	137.0	17.8	66	24.4
Autumn	138.3	18.7	63	23.3
			271	100.1

SPEED IN INTELLIGENCE

MILES A. TINKER
University of Minnesota

Since the earlier investigations on the measurement of intelligence there has been persistent inquiry concerning the relation of quickness of response to ability. The frequent appearance of both discussions and experimental papers on this subject reveals an unabated interest in the problem and indicates to some degree its real significance in the field of psychology and education. It is quite evident that workers in the field are not satisfied with the negative and conflicting results obtained in the various experiments and are seeking a clearer definition of the issues and more adequate data from which to draw conclusions.

The variety of methods employed in studying the relation of speed to intelligence have undoubtedly tended to produce the marked variability in the conclusions reached. From a survey of the literature one finds that many investigators apparently failed to consider certain implications involved in their procedures. The purpose of the present paper is to point out the limitations of certain techniques employed, the invalidity of some of the assumptions, and a new approach to the problem which should result in more valid findings.

When intelligence and achievement tests were first devised for group testing it was necessary to choose between two techniques for administering the tests: (1) the time-limit procedure in which the amount of work done in a given number of minutes or seconds is measured; and (2) the work-limit procedure in which is measured the time taken to complete a certain task.

Although Whipple[1] preferred the work-limit procedure, Paterson and Tinker[2] have showed that the two methods yield equivalent scores in situations where speed of response is strongly emphasized. In the group testing situation a certain amount of disorder and confusion is apt to arise when some members of the group finish early and then wait for the others to complete the task. Also the period taken for the slow workers to get through is often very time-consuming. Therefore, because of expediency, the time-limit procedure soon became the almost universal method of administering group tests. Since the time-limits allowed for the tests or parts of tests tended to be comparatively short (some as short as 1½ minutes) there has been inherent in the testing situation an emphasis on speed, even where the test itself has not been labelled "speed test." Even in the "power" (or ability, or all-or-none) tests, in which level of performance is emphasized, the factor of quickness of response has not been entirely eliminated. "The essence of the intelligent reaction seems to be the solution of a problem against time."[3] When a subject does solve a difficult problem, the time taken is important. This is emphasized particularly in the practical life situation and in scholastic performance where a speedy solution of problems distinguishes the efficient from the inefficient. As Peak and Boring have pointed out, the power or ability tasks are not tests of performance that do not involve speed but are merely tests which do not take account of speed.

It is very difficult to classify most intelligence examinations as either "speed" or "power" tests since they

[1] G. M. Whipple, *Manual of Mental and Physical Tests*, 1914, I, 8-9.

[2] D. G. Paterson and M. A. Tinker, "Time-Limit vs. Work-Limit Methods," *Amer. J. of Psychol.*, 1930, 42, 101-104.

[3] H. Peak and E. G. Boring, "The Factor of Speed in Intelligence," *J. Exper. Psychol.*, 1926, 9, 71-94.

usually contain both easy and more difficult problems. These easier problems have some degree of complexity and any task, easy or difficult, is time-consuming. As the complexity of the problems increases, level of performance is raised. It is impossible, therefore, to make a clear distinction between "speed" and "power" tests or parts of a test. Who can say where one ends and the other begins?

The time limits imposed in the testing situation vary from very strict to very liberal. However, it is almost impossible to find a test in which the examining situation, despite directions printed on the examination blank, does not set up the speed attitude, or in which the subjects could not improve their score in a time allowance that is really unlimited. Most time limits which are designed to permit the subject to reach his highest possible level of performance are far from being unlimited. It is probable, therefore, that in practically all group tests now used the testing situation sets up in the subject the speed attitude to some degree. Under these testing conditions the subject usually maintains a more consistent working pace than if he considered that rate of work would have no influence whatever upon his final score.

Since rate of work seems to be a determinant of intelligence test scores it is important to know the relation between speed and intelligence (or ability score). In general, four methods have been employed to study the factor of speed in intelligence. Let us consider the validity of each of these.

In the most persistently used procedure speed of response in reaction time or in some other simple motor test has been compared to intelligence as standardized tests measure it and as teachers judge it. These experimenters seem to have assumed that rate of response in reaction time or in the other motor tests used represented a general rate of response for the subject tested. This has been shown to be a false assumption for, in

several experiments where the interdependence of speed in various types of motor performance was determined, motor ability (speed of response) was found to be specific to a large degree. There is no general motor ability comparable to "general intelligence." In the light of the specificity of motor ability, it is no surprise that in well controlled and adequately planned experiments no appreciable relationship was discovered between speed of simple motor response (including reaction time) and intelligence. The negative results obtained in this type of experiment have served a good purpose in demonstrating the futility of seeking in a highly specific speed score a measure of general intelligence.

Since there are marked individual differences in reflex conduction rate it was pertinent to inquire whether these reflex times are basic speed measures which may be associated with intelligence. The most adequate investigation on the subject is reported by Travis and Young.[4] A correlation of practically zero was found between reflex time and intelligence and this was confirmed in other experiments. Since there is no evidence concerning the relation between reflex time and speed of nerve conduction within the central nervous system, experiments like Travis and Young's throw no unequivocal light on the possible relation of intelligence to speed of nerve conduction in complex mental operations.

A group of experiments, typified by that of Ruch and Koerth,[5] have investigated the influence of restricted time limits on test performance by correlating scores under standard time with scores under double or unlimited time. These correlations were very high and were

[4] L. E. Travis and C. W. Young, "The Relations of Electromyographically Measured Times in the Patellar and Achilles Reflexes to Certain Physical Measurements and to Intelligence," *J. Gen. Psych.*, 1930, 3, 374-400.

[5] G. M. Ruch and W. Koerth, "'Power' vs. 'Speed' in Army Alpha," *J. Educ. Psychol.*, 1923, 14, 193-208.

interpreted to mean that a subject's relative standing was not markedly affected by rigid time limits. Since other investigators[6] have found under certain testing conditions that the correlations between scores under standard and unlimited time deviate somewhat from unity, there are certain individuals who are handicapped by an imposed time limit which prevents them from attaining the relative levels of which they are capable. Although the authors have never claimed it, these high correlations between scores under restricted and extended time-limits have been interpreted by a few writers to signify a high relationship between speed of response and level of ability (altitude). This latter inference is, of course, unjustified with the evidence now available. The correlations merely indicate the degree to which relative level of attainment is affected by restricted time limits.

Another approach to the problem of speed in intelligence is typified by Hunsicker's[7] experiment. The times taken to do easy arithmetical problems and sentence completions were correlated with level attained in difficult arithmetical problems and sentence completions. The average correlations (corrected) between rate of work on the easy material with level of ability were .46 and .47 respectively for the arithmetic and completions. Hunsicker concluded that "there is a consistent and fair positive relationship between rate and ability." These conclusions and many others based on similar procedures are based on the assumption that rate of work on easy items is an approximately pure measure of speed, and that this rate of performance is the same for response on difficult items. There are no data to support this

[6] F. S. Freeman, "The Factor of Speed," *J. Gen. Psychol.*, 1932, 6, 462-468.

[7] L. M. Hunsicker, "A Study of the Relationship Between Rate and Ability," *Teachers College Contributions to Education*, No. 185, 1925, Pp. 52.

assumption. In fact the evidence indicates that measures of simple mental functions tend to correlate low with measures of complex mental functions. All that Hunsicker's data reveal is a small correlation between rate of work *on easy test items* and level of achievement *on difficult test items* which have similar content.

It seems to the writer that Hunsicker and several of the other experimenters dealing with speed in intelligence have failed to advance materially toward a solution of the problem of speed in intelligence because they have uniformly insisted that speed be measured by a rate test of "rare purity." Such rate tests are composed of easy elements which can be readily answered by all subjects and scored in terms of time taken to complete the tasks. Apparently the assumption made has been that such a measure yielded a rate of work which was common to all mental activities of the subject, or at least to the solution of the more difficult elements of content similar to that in rate test. There are abundant data available which show that simple, precisely measured functions produce scores which yield low correlations with each other and with complex functions. It is difficult, therefore, with the evidence now at hand to justify the assumption that speed on one of these simple tasks yields a measure which is the natural rate of work for this subject in the complex mental performances which result in an intelligence test score. The conclusion that the correlation between such a rate score and ability (level of performance) represents the relation between speed and intelligence is unconvincing. A different appriach to the whole problem may be helpful.

In considering speed in intelligence the practical, and probably also the most important consideration is how long it takes a subject to solve a problem irrespective of whether the task is easy or difficult. As stated above, intelligence as the tests measure it or in the natural

life situation is essentially the solution of a problem against time. The most logical way to measure speed in intelligence is to measure the time taken to solve the various elements constituting the intelligence test. It might be argued that such a procedure would not produce a "pure" measure of rate of work since other factors besides speed might affect the score, i. e., desire to get a high score, distractions, fatigue, dwaddling and the like. If this were true it would be difficult to motivate a consistent attitude or rate of work throughout the testing period. But why insist on a "pure" measure of rate of work if a subject's natural rate of work is partially conditioned by distractions and the like? It is still his natural rate of work and the one we are most concerned with. Furthermore there is much evidence which indicates that we can motivate a subject to maintain a consistent rate of work in test response. In the relatively simple speed tests the subject works with a speed set or attitude, and this set does produce a consistent rate of work. In the more difficult mental examinations with time limits which are restricted to a greater or less degree, the subject is working against time which involves consistent application of his energies to the task of problem solving. It has been argued that such motivation sets up a speed altitude which handicaps the subject. However, we have rather conclusive proof to the contrary in the experiments with standard and extended time. One's relative standing in a group is little affected by the strict time limits, even on the more difficult power tests such as the Ohio State University Psychological Test. Since it is possible to promote a consistent working attitude in tests with strict time limits, it should be equally easy to set up a similar attitude when the subject is working against time to complete a test item in the shortest interval possible. It seems reasonable, therefore, to maintain that the time taken to complete a test

item on a difficult level, and with a definite set for speed, yields a rate score which is characteristic of the subject for that type of task.

In view of the above inferences, it seems also reasonable to suggest that a more accurate picture of the relation between speed and intelligence would be obtained by measuring speed and level of ability on exactly the same test and correlating the two scores. To get an adequate picture of this relationship, tests of various levels of difficulty should be employed for it is probable that the relation of speed to intelligence varies somewhat with the nature of the test used. In fact, each level of difficulty may have a speed of response which is fairly specific to that level. A preliminary study with this approach to the problem is under way at Minnesota.

LEARNING TO GLIDE

ARTHUR S. OTIS
Yonkers, N. Y.

Gliding, as you know, refers to piloting a motorless airplane. A motorless airplane cannot move horizontally in still air at a uniform rate, as can a motor plane. It depends upon gravity to propel it and when it is flying in still air it must therefore sink slightly as it moves forward. However, gliders have now been constructed which can fly almost horizontally; that is, if properly piloted they can fly perhaps ten feet forward with a drop of only one foot. Hence a glider, if started at the top of a hill, can travel a long distance before finally coming to rest on the ground.

When a wind blows up a hill the upward motion of the air counteracts the downward motion due to gravity, and the plane can be kept aloft for several minutes or even hours. Thus, if the up wind is blowing northward, the pilot can glide east and west, back and forth, for an indefinite time. This is called sailing.

Piloting a glider is similar to piloting a motor plane; that is, the controls are so arranged that the steering of the two kinds of planes is practically the same.

Now it so happens that in a section of Germany called the Rhön there is a mountain, the rounded top of which is bare of trees and on one side of which there is a long, gradual, grassy slope of nearly a mile, with a large level spot at the foot of the slope for landing, making in all a wonderful place to glide. There is frequently a wind blowing up the hill, which permits sailing (segelfliegen). The place is called Wasserkuppe (water cap) because of the presence of springs almost at the summit of

the mountain. And here there is a school for gliding, said to be the best in Germany.

While taking a course in gliding at Wasserkuppe last summer I was surprised to discover the large number of elemental ideas and associations that go to make up the rather complex habits which must be formed before one can successfully pilot a plane.

Each primary motion of the plane is made of course by a specific motion of the hand or foot, and the association of the specific motion of hand or foot with the needed motion of the plane may be thought of as constituting an elemental idea in learning to fly. These elemental ideas are gradually combined into more and more complex habits until a pilot can execute a series of movements of a precise amount and so timed as to accomplish a very precise purpose in piloting the plane, and the execution of this series of movements seems to the pilot like just one thought.

Now it was of much interest to me to know just what elemental ideas were involved in learning to glide and how many of them there were. Hence between flights I analyzed the possible motions of the plane and made an inventory of these elemental ideas. I offer them herewith for what interest they may have for students of psychology.

The flight of a glider is started as follows. The pilot, with helmet on his head, is strapped into a seat in the front of the plane. The plane is held level (balanced on a keel) by the instructor. Two students sit on the ground and hold the tail of the plane. Two large rubber ropes are joined to a ring which is placed in a downward-pointing hook at the nose of the plane. Five students take hold of the forward end of each rope, the two ropes forming a V with the plane.

The pilot gives the commands "Achtung" (Everybody ready), "Ausziehen" (Pullers walk forward), "Laufen" (Pullers run), "Los" (Holders let go). The plane shoots

forward into the air, quickly attaining a speed of perhaps forty miles an hour. The ropes drop, and the pilot is off.

Needless to say the first flights are made from a very low hill and later from higher ones. The first goal to be achieved is a flight of thirty seconds. When this is accomplished one becomes an "A" pilot. Later one learns to make longer flights, culminating in sail-flying in which, with the help of upward currents of air, one can maintain altitude for many minutes or even hours.

Thirty seconds sounds like a short time, but in that time one traverses perhaps five or six hundred yards. In order to remain in the air for thirty seconds the student must have learned to make, more or less automatically, all the basic movements necessary to guiding the plane. Refinements such as flying with a side wind, turns with "banking," etc., come later.

There are three basic movements of a plane to be governed: (1) rotation on an axis from nose to tail, which let us call *balancing;* (2) regulation of the angle of descent by pointing the nose of the plane higher or lower, which let us call *pitching;* (3) changing the direction of the plane with reference to points of the compass by turning right or left, which let us call *steering.*

The pilot has his right hand on a "stick" which is held between the knees. The stick is normally held in a vertical position. He has his feet on a cross bar much like that on the front of a modern flexible sled. By moving the stick, he balances and pitches, and by moving the cross bar, he steers the plane.

The ideas involved in learning to direct a plane are as follows.

I. BALANCING

1. The plane rotates on a horizontal axis (front to back) movement governed by a sidewise motion of the stick which moves the ailerons (flaps of the wings).

2. When the stick is in a position tilted to the right, the plane rotates to the right (clockwise) and the same on the left (i. e., left tilt of stick produces left rotation).

3. The farther to the right (or left) the stick is tilted, the more rapid the rotation. It is as in steering an automobile—the farther the wheel is turned, the more rapid the turning of the car.

4. But motion of the stick is always to be thought of as relative to a line through the wings of the plane.
Consequently

5. When the stick is "straight" (perpendicular to the wings) there is no rotation—the plane maintains its position, even though tilted.

6. With (uniform) motion of the stick to the right from "straight" there is (uniform) increase of speed of rotation to the right, and the same on the left (i. e., with (uniform) motion of the stick to the left from "straight" there is (uniform) increase of speed of rotation to the left).

7. With (uniform) motion of the stick from the right back to "straight" there is (uniform) decrease of right-hand rotation until the plane assumes a fixed tilted position; and the same on the left.

8. To rotate the plane from level to a position tilted to the right, two motions of the stick are necessary—first a tilting of the stick to the right to produce rotation, then a return of the stick to "straight" to stop the rotation; and the same on the left.

9. Moreover, the rapidity with which the rotation is made is governed by the amount of tilt of the stick from straight.

10. Also, the return of the stick to straight must be so timed that it reaches straight at the exact moment that the desired position (tilt) of the plane is reached. If the stick is returned to straight too soon, the amount

of tilt of the plane will be less than desired, and if the stick is returned to straight too late, the amount of tilt of the plane will be greater than desired.

11. Similarly, to rotate the plane from a position tilted to the right, back to level or any desired amount in a counter-clockwise direction, two motions of the stick are necessary—first a tilting of the stick to the left from straight to produce a left-hand rotation, and a return of the stick to straight to stop the rotation at the desired point. (And the same on the other side.)

12. As before, the return of the stick to straight must be effected at the exact instant the level (or desired) position of the plane is reached.

13. If, in the attempt (in No. 10 or 12) to give the plane a certain tilt, the stick is returned to straight too soon and the desired amount of tilt is not reached, the same two motions of the stick are needed again to give the plane the necessary additional tilt. And, of course, if the additional amount of tilt is small, the tilting of the stick must be of less extent or duration or both than before.

14. And, conversely, if (in No. 10 or No. 12) the stick is not returned to straight soon enough and the plane is tilted too far, then two additional motions of the stick opposite to those of No. 13 are needed to reduce the tilt of the plane and of course the stick must eventually be gotten back to straight at the exact moment that the desired degree of tilt of the plane is reached.

15. Incidentally, the flyer determines whether or not the plane is level by looking at the horizon (not at the ground) and determines the desired amount of rotation necessary to maintain balance or for other purposes by watching the horizon (remembering of course that if he is tilted to the left—or the horizon seems tilted to the right—he needs to make a right-hand rotation). *Now,*

16. If the flyer thinks his stick is straight but finds the plane rotating away from the desired position (level or any given tilt), this indicates that the stick really is not straight; hence two movements are necessary. Thus, if the plane is rotating to the right the stick must first be moved far enough to the left (past straight) so that the right-hand rotation will be stopped and a left-hand rotation produced sufficient to bring the plane back to level. Then as the plane approaches level the stick must be brought back to straight—*but* straight is not as far back as the stick was at first, so a shorter movement must be made. Either the proper amount of movement must be guessed at and later corrected if necessary by two more moves or by very careful coordination and attention to the changing positions of the stick and of the plane, and the stick must be brought to a stop at center just as the plane reaches level.

17. There is a contrariness between No. 16 and the preceding ideas because in order to produce a desired right-hand rotation, the stick is pushed to the right, but if an undesired right-hand rotation exists, it must be corrected by a *left*-hand movement of the stick. It is not sufficient therefore merely to associate right-hand stick movement with right-hand rotation. Moreover, the correction of undesired rotation (maintaining balance) is more necessary at first than producing tilt.

18. Now as the learning process begins, these ideas are all more or less isolated and cannot be called to mind quickly enough to enable one to balance a plane successfully. Hence it is necessary eventually to learn to think of the two motions mentioned in No. 8 as a single idea—a single motion consisting of two parts, both of which are contemplated in advance and executed as a single reaction.

19. Also, Idea No. 9 must be eventually combined with No. 8 so that the pilot may think in advance not

only of the double motion he is to make but what extent of motion will be made.

20. Also, Idea No. 10 must be eventually combined with Ideas 8 and 9 so that with one thought, so to speak, the pilot can decide in advance to make a certain double motion of a certain extent and at the same time be on the alert that the duration of the movement is properly timed.

21. This combination of ideas must be associated of course with the proper stimulus so as to constitute a more or less automatic reaction because, when traveling ten, fifteen, or twenty feet above the ground, there is no time to "think." The motions must be made in the proper sequence, having the proper extent and the proper duration, or the plane may dash to the ground.

So much for the moment regarding balancing the plane. Let us now consider pitching.

II. PITCHING

(This means pointing the nose of the plane downward or upward, or, in other words, rotating it on an axis through the wings.)

1. All that has been said above regarding balancing applies similarly to pitching. That is,

(a) The nose of the plane is raised by a pull of the stick and lowered by a push. These motions actuate a horizontal rudder at the tail of the plane.

(b) The amount of pull from the central position determines the rate of upturn of the nose, the amount of push from the central position determines the rate of down turn of the nose.

(c) A given change in pitch of the nose is accomplished by two movements of the stick, one to start rotation and one to stop it.

(d) As before, the return of the stick to neutral must be accomplished at the exact instant the desired pitch is reached.

(e) The various ideas involved in one act of steering must be combined as before into a single whole and executed more or less automatically.

There are, however, several additional ideas involved in pitching that are not required for merely balancing.

2. Thus, in order for a plane to glide it must point slightly downward. Moreover, there is an optimum angle at which the machine will glide best and farthest. Now a beginner can tell very easily whether a plane is balanced (level), but the proper angle for gliding is quite an unknown quantity for a beginner. One must learn first therefore what the proper gliding angle of the plane is and how to attain it.

3. Now it is even more difficult than balancing to determine at what angle the plane *is* (to say nothing of the angle at which it *should be*). Thus, a slight tilt of the horizon away from level is very easy to detect, but a slight rise or fall of the horizon is not so easy to detect because there is no cross-hair or other "sight" to go by. The important additional problem then relating to pitching is the finding of the optimum gliding angle.

4. It should have been explained above, perhaps, that if a plane is pitched too high, it will lose forward speed, "stall," and dive. If pitched too low, it will strike the ground before the maximum distance has been covered. Probably the best place to learn the optimum gliding angle of a plane is on a long, even slope which is itself inclined at the proper angle. The procedure is then as follows.

5. First learn to keep your eyes on the horizon or some point in the distance far enough to maintain a fixed position in the field of vision. This is called the *"Blickpunkt"* (point of sight).

6. Learn to observe at the same time out of the corner of your eye, so to speak, whether the ground is receding or approaching the plane—in other words,

whether the plane is leaving the ground or approaching it.

7. Notice whether the *Blickpunkt* is rising or falling in the field of vision—in other words, whether the plane is pitching downward or upward. (If the nose of the plane is pitching downward, the *Blickpunkt* is rising, and vice versa.)

8. Notice also whether the *Blickpunkt* seems high or low in the field of vision. It is the proper position of the *Blickpunkt* in the field of vision that must be eventually learned.

9. Now if it is observed out of the corner of the eye that the plane is approaching the ground, this indicates that the plane is pitched (pointed) too far down. The pitch must be raised, therefore, by two motions of the stick—a pull to lift the nose and a push back of center to stop the lifting at the desired point.

10. While the lifting is being done, the *Blickpunkt* sinks. The amount of sinking must be noted, for when the plane has ceased to approach the ground the *Blickpunkt* is in the right position in the field of vision.

11. The attempt should then be made to keep it there by means of the motions mentioned in No. 1.

12. Similarly, of course, if it is observed that the plane is receding from the ground, it means the plane is pointed too high, and the two opposite motions of the stick are needed—first a push to pitch the plane downward, then a pull to center to stop the movement when the plane has reached the right angle. The *Blickpunkt* will rise meanwhile to the proper position in the field of vision.

13. This proper position of the *Blickpunkt* should be fixed in mind, for later on when gliding is done over ground that is uneven, the above procedure cannot be used, and the pilot must rely on his memory of the proper position of the *Blickpunkt* in the field of vision.

14. When flying over uneven ground the proper gliding angle is maintained by noting the position of the *Blickpunkt* in the field of vision. If it begins to sink, the plane is pitching upward and the stick is not on center. The stick must be brought to center to stop the upward movement of the nose. The plane must be brought back to the proper gliding angle by two additional motions of the stick, as explained in No. 12. All these motions, however, must be combined into one idea and executed as a unit.

15. Similarly, if the *Blickpunkt* is found to be rising, an opposite series of motions is necessary. In every case the stick must be brought back to center at the exact instant that the proper angle is reached.

16. Learning to pitch a plane, therefore, consists in forming the association of stimulus and reaction, the stimulus being the falling away of ground or the rising of the horizon or both, or the rising of ground or falling of the horizon or both, and the reaction being the proper set of motions of the stick in proper sequence of proper extent and duration.

III. STEERING THE PLANE

(Changing its course with reference to points of the compass.)

1. The plane is directed with the feet on a cross bar which actuates the vertical rudder at the tail of the plane.

2. A push with the right foot causes the plane to turn to the right (the *opposite* of what happens when steering a sled or bicycle). Equalizing the feet stops the turning. The same on the left.

3. The farther the foot is pushed, the more rapid the turning.

4. Hence, to effect a turn through a given angle two motions are necessary as before, one to start the turning and one to stop it—the same as steering an automobile.

5. The type of learning is the same, therefore, in steering as in balancing and pitching. The learning is delayed, however, by the fact that it is *unnatural*, and all tendency to natural reaction must be overcome and a new and unnatural reaction substituted.

V. Complications

1. Of the complications that are encountered in flying, one is the need to make two adjustments at once. For example, it may be necessary to correct both balance and pitch at the same time. This calls for diagonal motions of the stick, which need to be learned as separate ideas.

2. Another complication is the fact that balancing requires much more force and more amplitude of motion than pitching, because of the construction of the plane, so that the same amount of rotation on different axes calls for quite different amounts of muscular effort.

3. A third complication is the fact that the amount of motion of the stick or foot bar necessary to accomplish a given amount of rotation varies with air speed (speed with reference to the air)—when flying rapidly much less control movement is needed than otherwise. The amount of movement necessary under the varying conditions of air speed must therefore be learned.

4. A fourth complication is that the neutral position of the stick is different for different planes, and when learning to glide one is constantly changing from one plane to another, as it happens.

5. A fifth complication is caused by the varying velocity of the wind, such that when flying against a strong wind the speed with reference to the ground is much less than the speed with reference to the air. Hence, one must learn to gauge air speed by the singing of the wires and whistle of wind past the ears in order to determine whether or not the proper gliding angle is being maintained.

6. A sixth complication is introduced by side wind which causes the plane to move in a direction other than that in which it is pointing. One must learn, therefore, to recognize "drift" and to detect the true direction in which the plane is moving. Indeed the feeling of a drifting plane is quite different from that of a straight-forward-moving plane, and the balancing, pitching, and steering reactions must be more or less separately learned for a drifting plane.

7. A seventh complication is the need to "bank" (tilt) the plane when turning. A double motion of the stick to give the plane a required tilt is combined with a single motion of the foot bar to produce turn. Then a double motion of the stick brings the plane back to level balance while a return of the foot bar to normal stops the turn. The foot and hand work must be so coordinated that the plane does not "slip." (When a plane is properly banked on a turn, the pilot feels no sideward force of the turning. Physically speaking, the tangent of the angle of inclination must just equal the ratio of the centrifugal force of the turning to the weight of the machine and pilot.)

8. A complication that has not been mentioned at all is *fear*. Have you learned to ride a bicycle? Imagine yourself never having been on a bicycle. A helmet is put on your head, you are strapped to the bicycle, taken to the top of a hill, pointed down a road, given a sudden push at 40 miles an hour, and left to your own devices. Can you imagine the feeling! (That is the way the first flight feels.) If you saw yourself headed for the ditch, could you remember what you had been told, namely, to turn in the direction in which you are falling!

Fear and concentrated thinking are incompatible. Thanks to an ingenious construction of *Zögling* gliders, a crash that may wreck the plane will leave the pilot unhurt. I saw many a crash and not a single injury during the course. It is only after fear is banished that normal learning can take place.

RETENTION OF KNOWLEDGE ACQUIRED IN A COURSE IN GENERAL PSYCHOLOGY[1]

ALVIN C. EURICH
University of Minnesota

When one college course is listed as a prerequisite for another, the assumption is made that students remember the content of the first course and use this knowledge while they are pursuing more advanced studies. At the University of Minnesota a two-quarter course in psychology, generally taken during the sophomore year, is a prerequisite for educational psychology. In view of this requirement it is pertinent to raise two questions: (1) To what extent do students possess the knowledge they acquired in general psychology when they register for educational psychology; and (2) Is the degree of retention related to the subsequent achievement in educational psychology? To seek answers to these queries is the purpose of this investigation.

The tests used were constructed by the Department of Psychology[2] as final examinations for psychology 1 and 2. They were given as retests at the beginning of the fall quarter 1931 to students registered for educational psychology 55. Many of these students had received credit for psychology 1 during the fall quarter of the previous year and for psychology 2 during the winter quarter.

The various types of items in each test and the number of each type are as follows:

[1] The University Committee on Educational Research of which Dean M. E. Haggerty is chairman subsidized this investigation.

[2] These tests were made available through the courtesy of Dr. Richard M. Elliott, Chairman of the Department of Psychology.

Type	Psychology 1	Psychology 2
Single choice	46	45
Analogies	46	40
Wrong word	36	45
Completion	23	
Matching		20
Total	151	150

The completion or the recall type of question appears only in the examination for psychology 1, whereas the matching type is included only in psychology 2. The single word, analogies, and wrong word comprise parts of both examinations. The last named is a new type[3] consisting of a false statement which the examinee corrects by finding the word that makes it false and writing this word at the beginning of the statement. In a study conducted by Dr. Longstaff[4] the reliability coefficients for similar examinations used during the years 1928-29 have been shown to range from .86 to .93.

Even though the examination was given to over 200 students in October, 1931, but 99 of this number had taken it the previous December as the final examination in psychology 1. The distributions of scores for these cases are arranged in Table 1. The mean of the December scores is 90 whereas nine months later the mean is slightly more than 66. The difference of approximately 24 is greater than 19 times its probable error and can be described, therefore, as being unquestionably significant. In terms of standard deviations the drop in mean scores is approximately 1.7.

[3] Grace Holmes and Edna Heidbreder, "A Statistical Study of a New Type of Objective Examination Question," *Journal of Educational Research*, 24:286-292, 1931.

[4] H. P. Longstaff, "Analysis of Some Factors Conditioning Learning in General Psychology," *Journal of Applied Psychology*, 16:9-48, 131-166, 1932.

TABLE 1

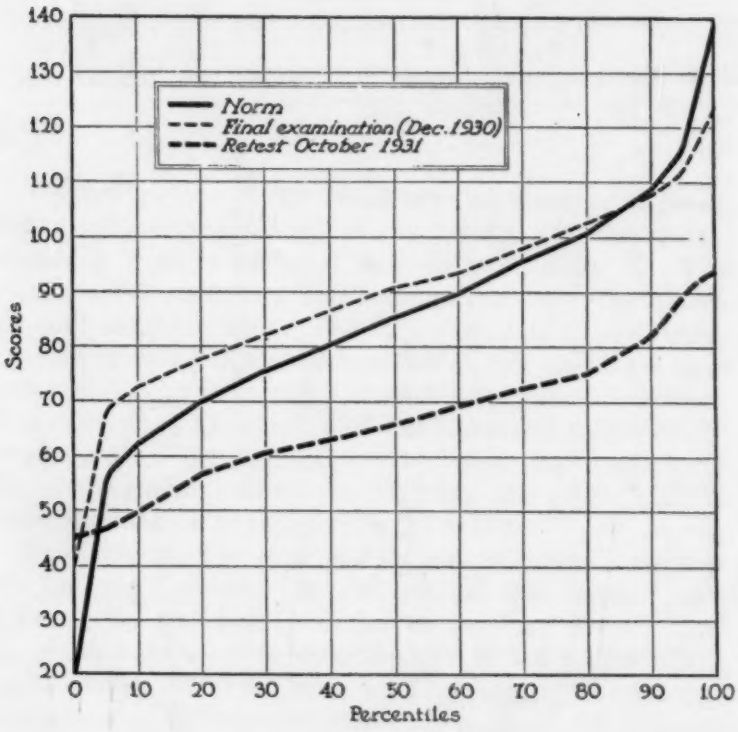
Distributions of Scores on the Final Examination in Psychology 1 and 2 at the Close of the Course and Six to Nine Months Later

Score	Psychology 1		Psychology 2	
	December, 1930	October, 1931	March, 1931	October, 1931
125			2	
120	1		5	2
115			11	2
110	6		12	3
105	7		12	11
100	11		13	9
95	11		7	11
90	17	7	6	12
85	10	1	8	14
80	12	3	5	11
75	9	8	1	3
70	9	18	1	
65	3	14		2
60	2	19		1
55		12		2
50		6		
45	1	10		
40		1		
N	99	99	83	83
M	90.03	66.29	104.01±.90	93.77±.98
D_{M1-M2}	23.74±1.23		10.24±1.33	
D/PE_D	19.30		7.70	
S.D.	13.65	12.00	12.10	13.20

The percentile curves in Figure 1 delineate more clearly the distributions of scores in October, 1931 and December, 1930. The uppermost curve represents the scores for the 99 students at the close of the course in psychology 1. The second curve is based upon the norms for 507 students who completed the course in December, 1930. The bottom curve pictures the distribution for the 99 students in October, 1931. It is readily apparent that

FIGURE 1

Percentile Scores on the Psychology 1 Examination at the Close of the Course and Nine Months Later



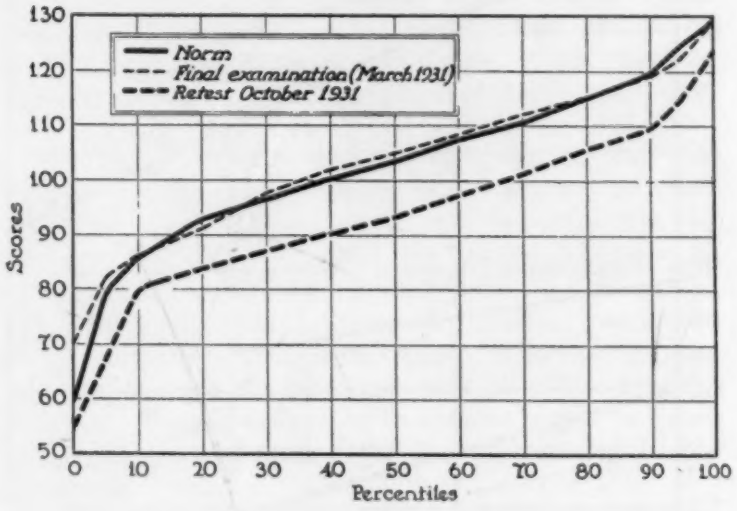
at the end of the general psychology course, the group retested has a median score approximately six points greater than the median for all students who took the course. At the upper end of the curves, however, the line depicting the scores for the entire group crosses and remains above the line for the retested group. After nine months the median score is at approximately the third percentile for the same group at the end of the course while the highest point of the curve does not extend much beyond the 60th percentile of the earlier dis-

tribution. Throughout, the distribution shows a definite loss over the nine months period.

Similar distributions for the examination covering the content of psychology 2 are arranged in Table 1 and in the form of percentile curves in Figure 2. These tests

FIGURE 2

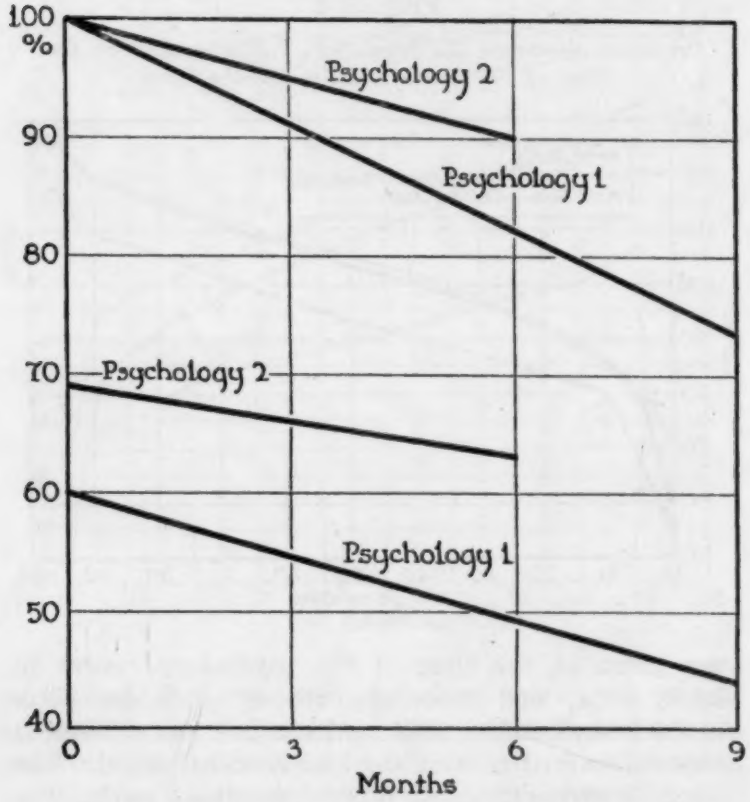
Percentile Scores on the Psychology 2 Examination at the Close of the Course and Six Months Later



were given at the close of the psychology course in March, 1931, and repeated with 83 individuals six months later, October, 1931. Although a loss obtains, it is not as marked as over the nine months period. The percentile curves illustrate this point rather clearly. The median score for October is at approximately the 23rd percentile rank as determined from the March scores. The highest score in October reaches the 94th percentile of the earlier distribution. In view of the fact that a summer vacation period of three months intervened, the loss does not appear to be very great. To be sure, it is

FIGURE 3

Curves of Forgetting in Psychology 1 and 2 Based Upon the Per Cent of the Average Score at the Close of the Course and the Per Cent of the Total Number of Items in the Tests



significant as the difference in means is 7.7 times its probable error.

The amount of retention for the two tests is presented in a different manner in Figure 3. The two lines at the top of the graph are based upon a percentage scale with

the mean score at the close of the course representing 100 per cent. On the psychology 1 examination the mean score nine months after the course was completed is approximately 73 per cent of the mean at the close of the course. On the psychology 2 examination, the mean score six months after the course is 90 per cent of the mean at the close of the course. As these curves are drawn the loss of information over a six months period appears to be greater on the psychology 1 examination than for the other test. In terms of the curves drawn from similar investigations the difference might be assumed to be even greater than shown on the graph as a greater loss would be expected during the early part of the nine months period than during the last three months. It is unlikely that the difference in the two curves is due to a dissimilitude in the amount retained over a six months period. It is more probable that the difference might be accounted for by the difference in examinations. It will be recalled that the test on the content of psychology 1 where the greater loss is observed contained completion items; whereas, on the other examination the questions were all of the recognition type. True, a difference in content might be listed among the possible causes for the discrepancy in the two curves.

The lower curves in Figure 3 have been obtained in a similar manner to those of other studies.[5] The mean scores are represented as per cents of total possible scores. The difference of nine per cent at the close of the course is probably due to the variation in the examinations. Similarly, the greater difference six months later may be ascribed to the nature of the examinations or the content covered. In other words, it may be stated that percentages representing the amount retained have

[5] Edward B. Greene, "The Retention of Information Learned in College Courses," *Journal of Educational Research*, 24:262-273, 1931.

meaning only with reference to the examination used. They cannot be regarded as representing the amount retained from a given course. For example, an examination could be constructed through a selection of items so that the per cent retained would be relatively small. Another examination could be devised on which the per cent retained would approach 100. Since both examinations could be based upon the content of a specific course, it would be difficult indeed to accept either of the results as indicative of the permanence of learning in that course. Again the difficulties of measurement obscure the real facts in regard to what students retain from their college courses. The retests in psychology, however, show that students retain a substantial amount of the *measured* knowledge six and nine months after they have taken the course.

Amount Retained Related to the Amount Learned

The relationships between the amount of information retained and the amount possessed at the end of the course are expressed in the coefficients of correlation between the scores obtained at these two periods. For the psychology 1 examination, the coefficient between the final scores and those nine months later was found to be $.59 \pm .04$. On the psychology 2 test, the correlation obtained between the final scores and those six months later is $.66 \pm .04$. While these coefficients are not strikingly high they represent a tendency for the students who received high scores at the close of the course to maintain similar positions six and nine months later. Also those who obtained relatively low scores on the first test might be expected to attain low ranks on a later testing.

The observed tendency, however, does not indicate whether the students with the highest scores retain relatively as well as absolutely more than the students with the lowest scores. To ascertain such a relationship a

formula derived by Harris[6] was applied. For the psychology 1 examination the correlation obtained by this process is $-.03 \pm .07$; for the psychology 2 test, $-.02 \pm .07$. An interpretation of these coefficients leads one to the conclusion that no tendency is apparent for students who earn the highest ranks at the end of the course to retain relatively more of this knowledge six and nine months later. These results are at variance with those obtained by Johnson[7] in a similar study of retention of botanical information.

Relation of Achievement in General Psychology to Final Marks in Educational Psychology

To determine the contribution of a background in general psychology to the study of educational psychology, the relationships were calculated between the achievement tests for the two courses. The correlations as arranged in Table 2 reveal a coefficient of .57 between scores on the examination at the end of psychology 1 and grades in educational psychology. Between the psychology 2 examination at the close of the course and educational psychology, the coeffcient of correlation is $.44 \pm .06$. When the retests are used to predict achievement in educational psychology, a coefficient of .39 for the psychology 1 and .48 for the psychology 2 examinations would yield some indication of the accuracy of prediction. While a tendency persists for those who obtain high scores on the psychology examinations to do superior work in educational psychology, the first distinction is not always followed by the second.

Undoubtedly the relationship between the examinations and achievement in educational psychology is de-

[6] J. Arthur Harris, "The Correlation Between a Variable and the Deviation of a Dependent Variable from its Probable Value," *Biometrika*, 6:438-443.

[7] Palmer O. Johnson, *Curricular Problems in Science at the College Level*, University of Minnesota Press, 1930, pp. 96-97.

termined in part by the fact that the students of superior ability do well in both courses, whereas those of lower ability receive the poorer grades. To obtain an adequate picture, therefore, of the contribution of success in general psychology to success in educational psychology, an attempt must be made to hold ability constant. In lieu of an experimental situation the method of partial correlation was applied with the third variable in each relationship being college ability as measured by the Miller Analogies Test. The relationship between this test and the marks in educational psychology is expressed by a coefficient of .49. The last column of coefficients in Table 2 shows the effect of "partialling out" intelligence as

TABLE 2

The Relation Between Scores on the General Psychology Examination and Final Marks in Educational Psychology

Marks in Educational Psychology and	N	r	P.E.	Partial Correlations-Intelligence Held Constant	P.E.
Psychology 1 examinations					
a. At the close of the course	98	.57	.05	.40	.06
b. Nine months later	98	.39	.06	.13	.07
Psychology 2 examination					
a. At the close of the course	82	.44	.06	.21	.07
b. Six months later	82	.48	.06	.26	.07

measured. It is evident that while a tendency is apparent aside from ability for students who achieve a certain level in psychology to maintain relatively the same level of achievement in educational psychology, the trend is so slight that it would hardly afford a basis upon which general psychology might be justified as a prerequisite for educational psychology. The justification must continue to rest upon logical assumptions.

SUMMARY

In this study an attempt has been made to measure the amount of general psychology that students retain at the beginning of a course in educational psychology and to determine the relationship between the amount retained and achievement in educational psychology. The accumulated results lead to the following general inferences.

1. After nine months the median score on a general psychology examination is at the third percentile of the same group as measured at the end of the course. Expressed in other terms, the mean score nine months after the course was completed, is approximately 75 per cent of the mean at the close of the course.
2. On another examination covering the content of the second quarter of psychology, the median score six months after the course was at the twenty-third percentile for the same group at the end of the course. The mean score on the retest is ninety per cent of the mean at the close of the course.
3. While relatively high correlations were obtained between scores on the final examinations and the retests, no tendency is apparent for students who retain absolutely more psychology than others to retain also relatively more of the knowledge they possessed at the close of the course.
4. When intelligence or college ability is held constant by means of partial correlations, the relationship between scores on the psychology examinations and achievement in educational psychology is positive but low.

AN EXPERIMENTAL STUDY OF CIGARETTE IDENTIFICATION

RICHARD W. HUSBAND and JANE GODFREY
University of Wisconsin

In a number of ways, through advertising and in personal conduct, there seems to exist an assumption that a person can identify his favorite brand of tobacco. The most striking example in advertising is the well-known "blindfold test" which has been run by one of the larger cigarette companies. The results—at least those which were published—all favored the company which conducted the tests. Persons were blindfolded and given several brands of cigarettes to smoke in succession, and were then asked to indicate their preference. It seems highly unlikely that the tests were run in an impartial, well-controlled manner.

This instance is only slightly more extreme than are cases which appear daily in magazines and newspapers. When Chesterfield advertises that it is milder, that it is compounded from an optimal blend of domestic and Turkish tobaccos; when Camels claim they use a more expensive raw product; when Luckies claim they use a toasting process—they state in substance that the facts or processes are such that the cigarette becomes preferable over others which are not made in that way.

As to personal conduct, the same implicit assumption appears at the consumer end of the economic order. The majority of cigarette users have definite preferences. Many individuals either refuse offers of other brands or smoke them with reduced satisfaction.

If both advertising statements and consumer preferences had factual bases, one would find that people could

discriminate different brands of cigarettes from each other. If tobaccos are so much alike that brands can not be told apart, or if the gustatory or olfactory senses are so dull that one can not spot his favorite brand, then we will be forced to admit that advertising and subjective preferences both are founded on hypotheses which are not based on facts or that suggestion is responsible.

It is, therefore, the purpose of our investigation to see whether persons, habitual users of one particular brand, can identify their preferred brand through taste alone.

TECHNIQUE

Fifty-one subjects were selected from a total group of nearly twice that size on the basis of questionnaire data. We only used persons who smoked regularly, at least several cigarettes a day; those who had smoked for at least a year and who preferred one particular brand. We did not demand that they never smoked another kind but that they would use just one make if it could be obtained with reasonable effort. We did not attempt to study pipe or cigar users. Actually, 21 subjects preferred Camels, 12 Lucky Strikes, and 11 Chesterfields, with a few other brands scattering. The three chief brands were used in all tests, with others added for subjects who used them, or for pragmatic purposes to keep supplies about even.

The subject was blindfolded and four cigarettes were placed on a tray. To ensure random selection and order he made his choice by groping in the dark. All he was told was that his own brand would be supplied along with three others. Several minutes were allowed between cigarettes until the subject felt there would be no interference. If he wished, he was given a mint to remove the taste but this usually caused more of an after-taste than a cigarette. The subjects were not compelled to smoke the whole cigarette, but could stop as soon as they had made their judgment. If they could

not name any brand, they were asked to indicate whether or not it was their favorite.

RESULTS

The main results are presented in Table 1. To see how often a brand was correctly identified as itself one can find the percentage quoted in both axes for any name. Obviously, 25 per cent accuracy should be

TABLE 1

Percentage Identification of Different Brands

	Camel Per Cent	Lucky Per Cent	Chesterfield Per Cent	20-Grand Per Cent	Spud Per Cent	Miscell. Per Cent
Camel	31	14	38	6	2	10
Lucky	19	41	21	4	0	14
Chesterfield	27	23	33	2	0	15
20-Grand	38	26	3	17	0	15
Spud	0	6	6	0	76	11

achieved through chance alone. We see that each brand is identified as itself in somewhat better than chance proportions. But the trend is not very pronounced, with Camels only 6 per cent better than chance, Luckies 16 per cent, and Chesterfields 8 per cent. 20-Grand, for some reason, has poorer than chance accuracy. We also find a few anomalies. Camels are identified as Chesterfields more often than as themselves. 20-Grands are said to be Camels more than twice as often as 20-Grands. Chesterfields were spotted correctly only a few times more than as two other brands. Spuds were identified correctly by practically all their users and by many others who had used them occasionally. Being mentholated, the proposition was somewhat different. It is as if a certain apple had a distinct taste of quinine; the merest novice would be able to identify it.

The average of cigarette users' correct decision in regard to their own brand was 31 per cent. This is very little better than chance.

Some rather interesting observations were made by the subjects during the tests. Many subjects expressed a liking for a particular cigarette and thought it was their favorite, and were terribly upset to find that it actually was one which they ordinarily refused to smoke, even if they had to go without. Others found cigarettes flat at times and judged them to be Chesterfields on the basis of mildness, while actually they were other brands as often, or more often, than that make.

SUMMARY

It is not our place here to condemn any advertising or to state that cigarettes do not have the qualities attributed to them. It is perfectly possible that Luckies are less irritating to the throat, that Chesterfields are really more mild, that a Camel does give a better taste, and that an Old Gold is really smoother. But it is certain that under controlled experimental conditions, subjects are unable to select their own brands from others on the basis of gustatory or olfactory cues alone.

THE USE OF VOCATIONAL AND PERSONALITY TESTS WITH THE DEAF*

VERNE WESLEY LYON

Illinois Institute for Juvenile Research, Chicago

The following studies were made at the Illinois School for the Deaf, at Jacksonville, Illinois, and pertain to the application of vocational and personality tests to deaf high school boys and girls. The average age of these students was nineteen years and the majority of them had been in school from five to ten years.

The first study is concerned with an objective measurement of the mechanical or motor ability of deaf boys, since it is believed that the best opportunities for the deaf are in those occupations which are concerned primarily with motor ability. The second is an attempt to determine the emotional or personality adjustment of deaf high school boys and girls, for an objective study of the emotional complications as a result of deafness is considered by those who have worked with deaf children to be of great value in their final understanding and ultimate vocational adjustment.

Fifty-nine boys constituting the 8th, 9th, 10th, 11th, and 12th grades at the Illinois School for the Deaf were the subjects for the measurement of mechanical, or motor, ability. The ages of these boys varied from thirteen to twenty-two years and the average was nineteen years.

The Minnesota Mechanical Test was selected as the medium for the measurement of this ability because of the high validity established on this test; the fact

* Studies from the Illinois Institute for Juvenile Research, Chicago. Series C, No. 233.

that it measures mechanical ability as distinct from intelligence and because of the simple directions that can be easily conveyed to, and comprehended by, the deaf.[1]

To facilitate the interpretation of results three dimensions are made in the distribution of percentile ratings. Group I ranges from 0 to 29, Group II ranges from 30 to 69, and Group III, from 70 to 100.

Interpretation of Results

Group I.—Twenty-six boys, or 44 per cent, of the group tested fall below the thirtieth percentile, which should indicate very inferior mechanical ability as measured by this test. This means simply that the efforts of these boys in occupational training should be concentrated more specifically on those vocations not requiring any mechanical aptitude.

Group II.—Twenty-one boys, or 36 per cent, of the group tested have what might be termed average mechanical ability. It is very difficult to make any definite suggestions with regard to this group without further individual study. They have a so-called general mechanical knowledge, but unless a boy definitely professes a liking for this particular field, it would be inadvisable to make any suggestions on the basis of this analysis.

Group III.—Twelve boys, or 20 per cent, of the group tested have high average or superior mechanical ability. For this group it can be said with some degree of assurance that the chances for these boys in work requiring definite mechanical knowledge and ability are very good. Opportunities for further training and development of their special ability should be made if at all possible,

[1] "It must be remembered that whatever these tests measure is mechanical ability, not mechanical capacity."

Paterson, Elliott, Anderson, Toops, and Heidbreder, "Minnesota Mechanical Ability Tests" (Minneapolis: *The University of Minnesota Press*, 1930), p. 7.

and if at the same time there is a definite interest or inclination along this line.

In summarizing the results of this test it is found that in the final analysis there are only two main groups to be considered, namely: those who do and those who do not have mechanical ability. The value of this classification is apparent. It eliminates the members of Group I from those vocations which are not adapted to their particular abilities. Further individual study of this group will result in a more complete understanding and proper vocational placement thus effecting a considerable saving in time, effort, and money to the school and a more valuable preparational training for the individual.

For the other groups where there is an indication

Percentile Rating of High School Boys at the Illinois School for the Deaf with Regard to Mechanical Ability as Measured by the Minnesota Mechanical Tests

Percentile Rating	No. of Students		
0-4	2		
5-9	3		
10-14	2	Group	26 Boys or 44%
15-19	8	I	
20-24	6		
25-29	5		
30-34	4		
35-39	2		
40-44	5	Group	21 Boys or 36%
45-49	3	II	
50-54	2		
55-59	3		
60-64	2		
65-69	0		
70-74	2		
75-79	1	Group	12 Boys or 20%
80-84	4	III	
85-89	5		
	59		

of mechanical ability the tests open up a large field of vocational possibilities, and at the same time present tangible information which should be of distinct value to a trained vocational adviser. It is highly probable that very definite vocational advice can be given to those who made a high score on this test. However, since the norms applied have been established on hearing individuals the evaluation of this classification will depend upon further investigation and application of results.

The Emotional or Personality Adjustment of Deaf High School Boys and Girls

It was deemed necessary, in order to give the best possible advice with regard to the vocational placement of the deaf, that some indication be secured with regard to their personality makeup. For this reason the Thurstone Personality Schedule was administered to eighty-seven high school boys and girls in order to gain some index of their neurotic tendencies. This was the first attempt, so far as the author was able to determine, that has been made to apply this questionnaire collectively to a group of deaf students. These questionnaires were presented in the presence of a psychologist to groups of fifteen to twenty students by an interpreter who was familiar with the instructions. Every precaution was taken to assure a complete understanding on the part of the students. Whenever an individual came across a word in the questionnaire, the meaning of which was unknown or obscure, proper meaning was given by the interpreter without suggesting an answer. The examiner felt that the co-operation and comprehension on the part of the students correspond closely with the usual testing situation for "hearing" individuals.

The general supposition among psychologists who have worked with this handicapped group is that the deaf present more emotional problems than do hearing indi-

viduals and if the results of this test are valid, these beliefs are substantiated.

The following table indicates the frequency distribution of total scores in the personality schedule.

TABLE I

Personality Score	Boys	Girls	Total	No.	%	Group
0- 4	0	0	0			
5- 9	1	0	1	1	1	A
10- 14	0	0	0			
15- 19	2	2	4			
20- 24	2	2	4	18	21	B
25- 29	6	4	10			
30- 34	6	2	8			
35- 39	2	1	3			
40- 44	4	1	5	42	48	C
45- 49	3	3	6			
50- 54	9	3	12			
55- 59	5	3	8			
60- 64	0	1	1			
65- 69	3	0	3	11	13	D
70- 74	4	1	5			
75- 79	1	1	2			
80- 84	2	1	3			
85- 89	2	1	3			
90- 94	2	0	2			
95- 99	0	0	0			
100-104	1	0	1			
105-109	0	1	1			
110-114	0	1	1	15	17	E
115-119	1	0	1			
120-124	1	0	1			
125-129	1	0	1			
130-134	0	0	0			
135-139	0	0	0			
140-144	0	0	0			
145-149	1	0	1			
Total	59	28	87			

Group	Personality Scores	Description
A	0-14	Extremely Well Adjusted
B	15-29	Well Adjusted
C	30-59	Average
D	60-79	Emotionally Maladjusted
E	80-	Should have Psychiatric Advice

It is noted that fifteen students, or seventeen per cent of the group measured, are classified in Group E as being "In need of psychiatric advice." Eleven students or thirteen per cent are classified in Group D as being "Emotionally maladjusted." The total number of students in both groups is twenty-six or thirty per cent who have indicated some degree of emotional instability. To quote, "We can be certain about one thing, namely, that those students who had a high numerical score, answering a large number of questions in the unfavorable manner, were certainly telling the truth about themselves."[2]

The large proportion of students classified in groups D and E may assume even greater significance when it is realized that the groups measured were decidedly homogeneous. That is to say, the students at the Illinois School for the Deaf are in direct association with each other for the greater part of the year and are therefore less likely to become isolated or unhappy because of uncontrolled environmental conditions.

It is expected that those deaf individuals who are not afforded the opportunity for this intimate contact would have a greater tendency to develop emotional problems.

In comparing the results of this study with that made on freshmen at the University of Chicago in 1929 it is found that the percentage of deaf students classified in Groups D and E is more than twice that of college freshmen of approximately the same age.

There are several factors which might cause one to question the validity of the results of this study on the deaf. First, the small number of cases on which to base results; second, the lack of understanding of words and situations by the deaf; third, the reliability of answers to an objective questionnaire; and fourth, the proper understanding on the part of the individual taking

[2] *Instructions for Using the Personality Schedule* (Chicago: The University of Chicago Press, 1929).

the test as to its purpose. It is impossible to control all these factors and the results must necessarily be considered as a preliminary investigation on which further experimental study should be made. This would involve comprehensive individual case study and the establishment of definite criteria for determining emotional adjustment.

THE PROGNOSTIC VALUE OF DIFFERENT TYPES OF TESTS IN COURSES IN EDUCATIONAL PSYCHOLOGY

PAUL W. TERRY
University of Alabama

INTRODUCTION

Within the last few years the necessity of giving greater attention in teacher-training institutions to the guidance and selection of young men and women who desire to teach has become apparent.[1] The fundamental and highly valued course in educational psychology provides one of the best opportunities for measuring the fitness of candidates.[2] Instructors of classes in educational psychology, moreover, need to know from which students much in the way of achievement is to be expected and those of whom little is to be expected. The profession, therefore, will welcome the development of instruments of measurement which promise to serve these purposes.

One of the recently offered instruments of this type is the Van Wagenen Reading Scales in Educational Psychology, Forms A and B, which were published in 1929 by the Educational Test Bureau of Minneapolis. The object of this study is to examine the availability of Form A of the Van Wagenen Scales for the purposes

[1] "The Direct Contribution of Educational Psychology to Teacher Training," *Yearbook* No. XX, National Society of College Teachers of Education, 1932. See Chap. II, "Selection of Entrants to Teacher Training Courses."

[2] *Ibid.* See Chap. III, "The Place of Educational Psychology in the Curriculum for the Education of Teachers."

above set forth by comparing its value in predicting achievement in educational psychology with the value of a test of general reading ability (Iowa Silent Reading Tests, Advanced Examination, Form A), and with the value of an intelligence test (Otis Group Intelligence Scale, Advanced Examination, Form A). Only seven of the ten tests in the Otis scale were used but the scores thus obtained were transmuted into ten-test scores.

The criteria of achievement included an objective test, an essay or traditional test, and the objective and essay tests combined in a single score. In the objective test were included 139 true-false, completion, and simple response statements that were concerned with the subject-matter discussed in the first half of the term. The essay test consisted of six questions calling for elaborate discussion. This test, which served as the final examination, was valued at 60 points. Corrected for attenuation the correlation of the essay with the objective test was .85.

The members of two classes in educational psychology, who were juniors for the most part, who were attending the summer school of the University of Alabama in 1932, and who were taught by different instructors, acted as subjects. It was intended at first to treat the data from the two classes separately. Their performances on the several tests, however, were so nearly the same that it was deemed best to treat them as a single group. In Table I the reader will find the number of students, the mean, standard deviation, range of scores, and the coefficient of reliability for each of the tests. The conclusiveness of the study is qualified by the smallness of the number of cases available. There is, nevertheless, a consistency in the findings which enables them to be presented as tentative conclusions with some assurance.

Scoring and Reliability of the Van Wagenen Scale

No indication as to the reliability of the Van Wagenen Scale is given in the accompanying directions and no publication concerning it has come to the writer's attention. It was necessary, therefore, to determine its reliability. The method prescribed for scoring the scale is somewhat complicated and it appeared difficult if not impossible to use this method in determining reliability by the "split-halves" process. It was not difficult, fortunately, to devise an alternative method of scoring that was suitable for this purpose. This method consisted essentially in treating the four or five statements under each of the fifteen paragraphs to be read as two-response items. Thirty-five of the sixty-three items were to be checked and twenty-eight were not to be checked. The score according to the new method, was the number checked or not checked correctly. It did not appear to be necessary to take account of guessing since there were from four to five statements under each paragraph and the number to be checked varied from one to four.

The soundness of this "number-of-rights" method of scoring was checked by correlating the scores thus obtained with the "C scores" which were obtained by following the directions accompanying the scale. The correlation resulting was .93. The probable error in predicting "C scores" from "number-of-rights" scores was found to be 1.96 points—a low error of prediction for a distribution the mean of which was 95.6 points (Table I). A very close relationship is indicated by these figures. In view of these facts it appeared reasonable to treat the "number-of-rights" scores as practically the equivalent of the "C scores"—as far as the purposes of this study were concerned. The "number-of-rights" scores were used, therefore, in all of the calculations reported below.

TABLE I

Data on the Several Tests

Data	Van Wagenen		Otis Intelligence	Iowa Reading	Objective and Essay	Objective	Essay.
	Rights	C Score					
No. of students	51	51	50	51	50	50	50
Mean	45.7	95.6	160.2	171.4	133.3	95.9	38.2
Standard deviation	6:5	8.0	26.6	49.0	21.6	17.1	8.3
Range of scores	29-59	73-111	88-205	82-282	50-172	41-127	9-55
Coefficient of reliability	.80		.96	.94*	.91	.93	.70

* For grade 12. *Examiner's Manual*, p. 19.

The split-halves technique of correlation when applied to the "number-of-rights" scores resulted in a coefficient of .67 between the odds and evens—which was "stepped up" to .80 by application of the Spearman-Brown formula. A reliability of .80 is fairly high, especially for a one-hour reading examination in a college subject of study in which the range of talent was limited practically to juniors. The reliability (.80) of the Van Wagenen scale compares favorably with that (.78±.018) reported by Eurich[3] for Part II (questions concerned with the comprehension of paragraphs) of the Minnesota Reading Examination for College Students, Revised Form, and likewise with the reliability (.79±.03) of Form A of the Minnesota Reading Examination XI and XIII (paragraphs with multiple choice questions) reported by Eurich on page 43 of the same monograph.

[3] Eurich, A. C., *The Reading Abilities of College Students*," University of Minnesota Press. (1931), p. 23.

Reliability of the Otis Group Scale of Intelligence

Before proceeding further it was necessary to ascertain the reliability of the Otis Group Intelligence Scale. The reliability of this scale with college students is not given in the manual of directions or in any of the literature which has come to the writer's attention. The 160 items of the 7-test scale were divided into evens and odds—there being 78 of the former and 82 of the latter. Correlation of the even and odd scores resulted in a coefficient of .931 which, when stepped up with the use of the Spearman-Brown formula, became .964 (Table I). This fact indicates that the Otis Scale was a highly reliable measure of intelligence with these college students.

Relation of the Van Wagenen to the Iowa and Otis Tests

Advantage should be taken here of the opportunity provided by the data in Table II to point out certain relationships between the specialized reading ability that is measured by the Van Wagenen Reading Scales in Educational Psychology, the general reading ability that is measured by the Iowa test, and the general intelligence measured by the Otis. Reading ability in educational psychology (Van Wagenen) is related less closely ($r = .63$)[4] to intelligence (Otis) than is general reading ability ($r = .75$) (Iowa), but both types of reading ability (Van Wagenen and Iowa) appear to be more closely related to intelligence (Otis) than they are to each other ($r = .51$). The Van Wagenen correlation (.63) with the Otis, however, is less than that ($.72 \pm .02$) of the Minnesota Reading Examination with the Minnesota College Ability Test as reported by Eurich[5] for 250

[4] All the correlations presented in Table II were corrected for attenuation.

[5] *Ibid*, p. 28.

freshmen. These facts imply that the Van Wagenen is less dependent upon the factor of intelligence than either of the two tests of general reading ability above mentioned.

TABLE II

Coefficients of Correlation
(Corrected for Attenuation)

Test	Van Wagenen	Iowa Reading	Otis Intelligence	Criterion Tests		
				Objective and Essay	Objective	Essay
Van Wagenen Reading		.51	.63	.72	.63	.77
Iowa Reading	.51		.75	.63	.58	.63
Otis Intelligence	.63	.75		.69	.70	.70

The greater independence of intelligence on the part of the Van Wagenen Scale as compared with the Iowa test is emphasized by the partial coefficients of correlation which are presented in Table III. When the Otis factor is parcelled out the correlations of the Van Wagenen with the objective and essay tests combined, with the objective test, and with the essay test, separately, are .51, .34 and .57, respectively. These partial coefficients are smaller by twenty points approximately than the correlations (.72, .63, and .77, respectively. Table II) of the Van Wagenen with the same three criteria. When the Otis factor is held constant the correlations of the Iowa test (.23, .12, and .22, respectively) with the three criteria are reduced by forty points approximately below the corresponding correlations (.63, .58, and .63, respectively. Table II) of the Iowa with the three criteria. Possibly it is to be expected that tests of reading ability

in a single subject, on account of this limitation itself if for no other reason, will prove to be less closely related to intelligence tests than tests of general reading ability.

TABLE III

Partial Coefficients of Correlation

First Test	Second Test			Third Test
	Objective and Essay	Objective	Essay	
Van Wagenen—and	.51	.34	.57	With Otis, constant
Iowa—and	.23	.12	.22	With Otis, constant
Otis—and	.44	.50	.43	With Van Wagenen, constant

THE VAN WAGENEN AND THE PREDICTION OF ACHIEVEMENT

The value of the Van Wagenen Scales to the profession will depend largely on the extent to which they predict achievement in educational psychology. In Table II it may be seen that the correlation (.72) of the Van Wagenen with the objective and essay tests combined in a single score is significant in magnitude—a judgment which a study of the scatter-diagram confirmed. A student's score on the two achievement tests combined can be predicted from his Van Wagenen score within a probable error of ±9.95 points. The mean of the distribution of the former scores is 133.3. In this connection the reader should be reminded that the average correlation between achievement scores in a single subject and intelligence scores seldom reaches the figure of .72.

The correlation (.63) of the Iowa test (general reading ability) with the objective and essay tests combined in a single score (Table II) appears to be less significant than that (.72) of the Van Wagenen (reading ability in educational psychology). A still lower correlation (.47±.08) between an objective achievement examination in educational psychology and a test of general reading ability, the Minnesota Reading Examination for College Students, Part II, was reported in a study of the records of 45 juniors.[6] Correlations of .42±.06 and .34±.06 between an achievement test in educational psychology and Forms A and B of another general reading test, the Minnesota Reading Examination XI XIII were reported by Eurich in the same investigation.[7] These facts taken together indicate that there is a considerable degree of validity in the Van Wagenen Scale and that it may provide a better means of predicting achievement in educational psychology than tests of general reading ability.

If subsequent studies yield correlations with achievement of similar magnitude the Van Wagenen Scale must be recognized as an instrument capable of yielding information that is valuable in giving guidance to prospective teachers. Instructors of classes in educational psychology with the aid of this scale may be able to determine with some degree of precision which of their students are making decidedly higher or lower achievement scores than their reading ability in the subject warrants and which students read so poorly that little achievement is reasonably to be expected of them. Those who are charged with the responsibility of counselling teachers in training may find the Van Wagenen scores helpful along with intelligence scores and other data used in guidance work.

The criterion of achievement above referred to was an essay test combined in a single score. At this point

[6] Eurich, p. 26.
[7] *Ibid.*, p. 44.

it is worth while to raise the question as to whether the predictive value of the Van Wagenen Scale is higher for one of these types of achievement test than for the other. Comparing the correlation of the Van Wagenen with the objective test (.63) and its correlation (.77) with the essay test, (Table II) indicates that the predictive value of the Van Wagenen is higher for tests of the essay type than for those of the objective type. The same observation though less marked follows comparison of the correlations (.58 and .63) of the Iowa with the objective and essay tests. Whether reading tests in general yield higher correlations with essay tests than with objective tests is a question well worth further investigation.

The Otis Test and the Prediction of Achievement

Data for comparing the Van Wagenen and the Iowa reading tests with the Otis Intelligence test in respect to the extent to which they predict achievement in educational psychology are also to be found in Table II. The Otis correlation of .69 with the objective and essay tests combined may not be significantly different in magnitude from the Van Wagenen correlation of .72 or the Iowa correlation of .63 with the same criterion. Unlike the reading tests, however, the intelligence test appears to be as closely related (.70) to the objective test as to the essay test. This relationship may not be as close as that (.77) of the Van Wagenen to the essay test. The closeness of the relationship of the Otis to both types of achievement test, as is indicated by the correlation of .70, however, serves as a warning that the value of the Otis as a means of predicting performance in educational psychology is not to be minimized. To this consideration in favor of the intelligence test is to be added the likelihood that intelligence scores are more closely related to achievement in other subjects than the scores on a test of reading ability specifically adapted to a single subject.

TABLE IV

Multiple Coefficients of Correlation

Tests Combined	Criterion		
	Objective and Essay	Objective	Essay
Van Wagenen and Iowa—with	.78	.72	.83
Van Wagenen and Otis—with	.78	.74	.82
Iowa and Otis—with	.71	.71	.72

Combinations of the Tests and Achievement

Data on the relationship of combinations of the Van Wagenen, Iowa, and Otis to the criteria of achievement are given in Table IV. The multiple correlations of the Van Wagenen and Iowa combined with the objective and essay tests combined in a single score (.78), with the objective test (.72), and with the essay test (.83), separately, are higher than the correlations (.72, .63, and .77, respectively) of the Van Wagenen alone with the same three criteria of achievement in Table II. Likewise, the multiple correlations of the Van Wagenen and the Otis combined, with each of the three achievement criteria (.78, .74, and .82, respectively), are higher than the correlations of either the Van Wagenen or the Otis alone with the achievement tests in Table II. The superiority of the multiple correlations of the Van Wagenen combined with either the Iowa or with the Otis is perhaps great enough to be taken into account by experimental investigators and by guidance officers. The situation of the instructor in educational psychology, on the other hand, with neither time nor money to spare for testing that can be dispensed with is different. His practical needs it appears may be met to some extent at least by the use of either the Van Wagenen or the Otis alone.

RELIABILITY, CONSTANCY, AND VALIDITY OF THE OTIS IQ

ARTHUR E. TRAXLER
University of Chicago High School

Most high schools now administer a group intelligence test to their pupils at or near the time of entrance to high school. In many schools the pupils are given only one such test and the IQ derived from the individual's score is used as an index of his ability throughout his high-school career. Since this is true, the question of the reliability, constancy, and validity of group-test IQ's becomes an important one. Unless the IQ secured from a group test is highly dependable, marked injustice may be done in the classification and grouping of individual pupils because of false information relative to mental ability.

During the past fifteen years the constancy of the IQ based upon the most widely used *individual* mental test—The Stanford Revision of the Binet-Simon Test—has been the subject of extensive investigation. As early as 1921 the studies in this field were sufficiently numerous that Rugg and Colloton summarized a number of the more important ones.[1] This article and later articles by other authors[2] indicate that the average variation in

[1] Harold Rugg and Cecile Colloton, "Constancy of the Stanford-Binet IQ as Shown by Retests," *Journal of Educational Psychology*, XII (September, 1921), 315-22.

[2] E. A. Lincoln, "The Constancy of Intelligence Quotients," *Journal of Educational Psychology*, XIII (November, 1922), 484-496; M. W. Wentworth, *Individual Difference in the Intelligence of School Children*, p. 27. *Harvard University Press*, 1926; Eustace Broom, "Constancy of the IQ," *School and Society*, XXV (March, 1927), 295-6; F. B. Randall, "Study of the Constancy of the IQ," *School and Society*, XXVI (September, 1927), 311-12.

Stanford-Binet IQ's is five or six points and that at least 20 per cent of the changes are greater than ten points.

Studies of the dependability of group-test IQ's are less numerous than those dealing with the Binet test. Two recent ones may be cited. Porter and Lauderbach report a study in which the Pintner-Cunningham Primary Mental Test was administered to a group of young children.[3] A small gain in IQ from the first to the second test was found. The reliability coefficient for 193 cases was .718. The authors thought that the rather low coefficient of reliability was due partly to the fact that kindergarten and primary children frequently lack a common fund of experience. They anticipated greater reliability in mental measurements in intermediate and upper grades because children in the higher grades have acquired common experiences through the environment of the school.

Lincoln and Wadleigh administered the Otis Primary, National A, and Terman A tests to 154 pupils in one of the grades of an elementary school in Reading, Massachusetts. Changes in IQ from test to test were studied. The median change was about 7.3 points and 36.9 per cent of the changes were ten points or more.[4]

It appears that these studies at the elementary-school level might well be supplemented with an investigation of the group-test IQ's of high-school pupils. One of the most widely used group intelligence tests at the high-school level is the Otis Self-Administering Test of Mental Ability, Higher Examination. For several years this test has been administered to pupils who enter the University of Chicago High School and during a period of four years it was administered to all the pupils enrolled

[3] M. J. Porter and J. C. Lauderbach, "On the Constancy of the IQ," *School and Society*, XXXIII (May, 1931), 675.6.

[4] E. A. Lincoln and Verna L. Wadleigh, "Change and Constancy in Group Test IQ's," *School and Society*, XXXV (April 30, 1932), 606-8.

in the school. An unusual opportunity is therefore presented for a study of the reliability, constancy, and validity of the IQ's secured with this well-known group test. Such a study will be reported in the rest of this article.

Reliability of Otis IQ's

The Otis Self-Administering Test of Mental Ability, Higher Examination, Form A, was administered to all pupils in the University High School in 1927. The same form was repeated in 1928. Form B was given in 1929 and Form C was used in 1930. The class graduating in 1931 participated in four administrations of the test, the class of 1932 took the forms given in 1928, 1929, and 1930, and the class graduating in 1933 took the last two forms. The intercorrelations among the scores made by the pupils in each of the classes are shown in Table I. The scores of pupils who missed one or more of the forms taken by their classmates have been eliminated from the calculations.

TABLE I

Reliability of IQ's Based on Otis Self-Administering Test of Mental Ability, Higher Examination, When the Forms of the Test are Administered at Intervals of One Year

Forms Correlated and Dates Administered	85 pupils graduating in 1931	100 pupils graduating in 1932	75 pupils graduating in 1933
Form A, 1927, with Form A, 1928	.710±.036	---	---
Form A, 1927, with Form B, 1929	.673±.040	---	---
Form A, 1927, with Form C, 1930	.676±.039	---	---
Form A, 1928, with Form B, 1929	.807±.025	.795±.025	---
Form A, 1928, with Form C, 1930	.721±.035	.647±.039	---
Form B, 1929, with Form C, 1930	.712±.036	.758±.029	.755±.033

The reliability of intelligence quotients derived from tests administered at least a year apart is, in all probability, lower than it would be if the tests were given with

only a short time interval between them. This situation is to be expected unless the IQ is perfectly constant. However, in view of the fact that the IQ's found for a class at the time of entrance to high school are frequently recorded and used for several years by the school, the correlation between the forms of a test administered in successive years is a matter of considerable practical importance.

Table I shows that the correlations between the IQ's from various forms of the Otis test range from .647 to .807. The mean of the ten correlation coefficients (.725) is not very high for a coefficient of reliability. It is not significantly larger than the reliability coefficient of .718 found by Porter and Lauderbach for the Pintner-Cunningham Primary Mental Test.[5] This study, therefore, fails to support their theory that greater reliability of mental measurements may be anticipated in the upper grades than in the lower ones. The coefficient falls within the category that Ruch has termed rather low.[6]

Although the reliability of the IQ found with the Otis Self-Administering Test of Mental Ability is not very high, it is only slightly lower than the reliability of the IQ secured with the individual Stanford Revision of the Binet-Simon Tests. The reliability of the Binet IQ's of 129 University Laboratory School pupils who were retested after intervals of one to four years is .762±.025. This correlation coefficient is only .037 higher than the mean of the Otis reliability coefficients. It appears that, although there should not be unquestioning faith in the reliability of the IQ's found with a single intelligence test —either group or individual—nearly as much dependence may be placed upon the reliability of the group-test IQ's as upon those of the individual test.

[5] Porter and Lauderbach, *op cit.*, p. 676.

[6] G. M. Ruch, *The Objective or New Type Examination*, p. 434. Chicago: Scott, Foresman and Company, 1929.

Constancy of Otis IQ's

Constancy of the IQ means the degree to which the IQ remains unchanged. A perfectly constant IQ would be exactly the same year after year. It is apparent, of course, that even if the IQ were constant, it would not appear to be entirely so when measured by an intelligence test, because the performance of each individual is influenced by a multitude of environmental factors which cannot be entirely controlled even under the best of testing conditions. The measured IQ will, therefore, fluctuate more than the real IQ. This article is concerned only with the constancy of the IQ of high-school pupils as measured by the Otis Self-Administering Test of Mental Ability.

Eight hundred eighty-five changes in the IQ's of University High School pupils were found and the means of the changes were computed. The IQ's were the ones that

TABLE II

Mean Change in Otis IQ of University High School Pupils

Groups and Test Forms	Mean Change
Class of 1931	
Form A, 1927, and Form A, 1928	5.6
Form A, 1927, and Form B, 1929	5.6
Form A, 1927, and Form C, 1930	4.6
Form A, 1928, and Form B, 1929	5.4
Form A, 1928, and Form C, 1930	4.7
Form B, 1929, and Form C, 1930	5.0
Class of 1932	
Form A, 1928, and Form B, 1929	5.5
Form A, 1928, and Form C, 1930	5.9
Form B, 1929, and Form C, 1930	6.0
Class of 1933	
Form B, 1929, and Form C, 1930	7.2
Mean of all changes	5.6

were utilized in computing the reliability coefficients reported in the preceding section. The means of the changes are shown in Table II.

The means of the changes in IQ range from 4.6 points to 7.2 points. It was stated earlier in this article that several studies have indicated that the average variation in Stanford-Binet IQ's is five to six points. The mean of the changes in Otis IQ (5.6) is about the same as the change in Stanford-Binet IQ.

The medians and quartiles of the changes in Otis IQ, gains and losses computed separately, are shown in Table III.

TABLE III

Medians and Quartiles of Changes in the Otis IQ of University High School Pupils

	Q1	Median	Q3
Gains	2.5	5.1	9.2
Losses	2.6	5.2	8.7
Both	2.5	5.1	8.9

The median gain and the median loss in IQ are slightly greater than five points. One-fourth of the changes in IQ are trifling in amount, being 2.5 or less. The highest one-fourth of the changes range from about nine points upward. The largest of the 885 changes was 25 points. In 38 cases, the change was zero—that is, the IQ on the repeated test was the same as the IQ on the previous one. In 155 cases, or 17.9 per cent, the change in IQ was 10 points or more.

It was pointed out near the beginning of this article that previous studies indicate that 20 per cent, or more, of the changes in Stanford-Binet IQ are 10 points or

more. It was also stated that Lincoln and Wadleigh found in a recent study that the median change among IQ's derived from the Otis Primary, National A, and Terman A tests was 7.3 points and that 36.9 per cent of the changes were 10 points or more. Compared with these figures, the changes in Otis IQ are small. Nevertheless, about one Otis IQ in six changes so greatly that marked inaccuracies might occur if one test alone were used for purposes of classification and grouping in high school. This is in accordance with one of the findings reported several years ago by Breed and Breslich in a study of the Chicago, Otis, and Terman group tests.[7]

A part of the change in IQ may be due to the practice effect of taking the test and to slight differences in the difficulty of the forms of the test. Some information relative to this point is presented in Table IV, which shows the mean scores made on the various forms of the Otis test.

TABLE IV

Mean Scores Made by High-School Pupils on Three Forms of the Otis Self-Administering Test of Mental Ability

Form and Date Given	Class of 1931	Class of 1932	Class of 1933
Form A, 1927	116.3	---	---
Form A, 1928	120.1	116.8	---
Form B, 1929	116.0	112.4	110.1
Form C, 1930	118.1	117.1	116.1

Some evidence of practice effect is found in Table IV. The mean gain from the first administration of Form A to the second administration, for the class of 1931, is 3.8

[7] F. S. Breed and E. R. Breslich, "Intelligence Tests and the Classification of Pupils," *School Review*, XXX (January and March, 1922), pp. 51-66 and 210-226.

points, and the gain from the first administration of Form A to Form C for the same class is 1.8 points. There is also a slight gain from Form A to Form C for the class of 1932.

Form B is apparently harder than Form A or Form C. It seems harder than Form A because, despite practice effect, there is a decrease in average IQ from Form A to Form B for the two classes who took both forms. It appears to be more difficult than Form C because, while the gain from the first to the second administrations of Form A, for the class of 1931, was 3.8 points, the gain from Form B to Form C, for the class of 1933, was 6 points. In other words, the gain from one form to another one not taken before was distinctly greater than the gain from the first to the second administration of the same form. This is contrary to what is to be expected if the forms are equally difficult.[8]

The difference in the difficulty of the forms and the effect of practice, although they cause the IQ to fluctuate, are constant factors which tend to influence the IQ's of all pupils alike. In so far as they cause identical changes in IQ from pupil to pupil, the changes are of slight importance. The variation in IQ already noted is, therefore, somewhat less significant than the data presented in this section suggest it to be.

Validity of Otis IQ's

The method used in standardizing the Otis Self-Administering Test of Mental Ability is intended to produce IQ's that are equivalent to Stanford-Binet IQ's. Stanford-Binet IQ's may, therefore, be used as a standard in studying the validity of Otis IQ's. The correlation between Otis IQ's and Stanford-Binet IQ's is one means of expressing their relationship. The correlation of Otis

[8] The data reported in this study relative to the differences in the difficulty of the forms of the Otis test are in agreement with the finding of Otis as stated in his manual of directions for the test.

and Binet IQ's for three groups of Sub-Freshmen in the University High School are shown in Table V.

TABLE V

Correlation of the Otis IQ's and Binet IQ's of Three Groups of High-School Sub-Freshmen

Class	Number of Cases	r P.E.
Sub-Freshmen, 1930-31	50	.799±.034
Sub-Freshmen, 1931-32	60	.732±.041
Sub-Freshmen, 1932-33	70	.622±.049
Mean		.718

The correlations shown in Table V are rather low for two tests designed to measure the same thing. The mean of the three coefficients (.718) is slightly lower than the mean of the intercorrelations between the various forms of the Otis test (.725). However, the relatively low correlation between the two tests is due mainly to the fact that the reliability of the IQ's is not very high. The correlation that would exist if the tests were perfectly reliable was found by applying Spearman's correction formula to the observed coefficient.[9] The correct correlation is .967, which is very high. The two tests would apparently rank pupils in IQ in nearly the same order if

[9] Spearman's formula for correcting for attenuation is

$$r_{tt} = r_{12}/\sqrt{r_{11}r_{22}}$$

where r_{tt} is the desired correlation between "true" scores freed from response error, r_{12} is the observed correlation, r_{11} is the reliability of test 1, and r_{22} is the reliability of test 2. (See Karl J. Holzinger and Blythe C. Mitchell, *Exercise Manual in Statistics*, pp. 112-113. Boston: Ginn and Company, 1929.)

they were perfectly reliable. From the standpoint of correlation with the Stanford-Binet test, therefore, the Otis test has satisfactory validity.

Although the correlation between Otis IQ's and Binet IQ's is as high as could be expected, the following table shows that Otis IQ's are not equal to Binet IQ's in magnitude.

TABLE VI

Mean Otis IQ and Binet IQ of High-School Sub-Freshmen

Class	Otis IQ	Binet IQ	Difference
50 Sub-Freshmen in 1930-31	111.0	118.9	7.9
60 Sub-Freshmen in 1931-32	108.3	118.0	9.7
70 Sub-Freshmen in 1932-33	114.8	121.2	6.4
180 Sub-Freshmen in 1930-33	111.4	119.5	8.1

In all three classes the mean Binet IQ is consistently higher than the mean Otis IQ. It appears that, on the average, about eight points should be added to Otis IQ's to make them equivalent to Binet IQ's.

SUMMARY

1. The Otis IQ is not high in reliability. The mean of ten reliability coefficients based on forms of the test administered at least a year apart was .725. However, the Otis IQ compares favorably in reliability with IQ's from other tests.

2. The Otis IQ is fairly constant, the median change from one form of the test to another being 5.1 points. About 18 per cent of the changes are 10 points or more. The evidence indicates that the Otis IQ is at least as con-

stant as the Stanford-Binet IQ and perhaps it is slightly more so.

3. From the standpoint of correlation with Stanford-Binet IQ the Otis IQ is highly valid. The correlation found in this study was .967 (correlation coefficient corrected for attenuation). However, the Otis IQ is, on the average, about eight points lower than the Stanford-Binet IQ.

4. It may be concluded that the IQ derived from the Otis Self-Administering Test of Mental Ability, Higher Examination, is about as dependable as the IQ secured through individual testing with the Stanford Revision of the Binet-Simon Tests. Teachers should remember, however, that in a considerable proportion of the cases, large errors in IQ occur and that the results of neither this test nor any other intelligence test can be accepted uncritically. Where possible, Otis IQ's should be verified through a repetition of the test or the use of some other intelligence test.

A TEST OF MENTAL INSTABILITY*

D. J. INGLE
University of Minnesota

INTRODUCTION

Among attempts made to develop group tests of mental instability, several paper and pencil tests of the questionnaire type are on the market. The items which make up these tests are questions each of which concerns a "symptom" of mental instability. It is claimed that individuals who respond with many atypical answers tend to be mentally unstable, whereas individuals who respond with few atypical answers tend to be mentally stable. Such an hypothesis involves the following assumptions: (1) that the questions concern "symptoms" which actually occur with a much higher rate of frequency in unstable individuals than in stable ones, (2) that the degree of instability increases as the number of "symptoms" increase, (3) that the meaning of each question will be interpreted in the same way by each subject, and (4) that each subject will give truthful answers. A critical review of the literature gives sufficient evidence that no paper and pencil test of mental instability has yet been adequately validated. Further attempts, then, to construct and validate such a test cannot be discouraged on the grounds that completely satisfactory tests are already available.

The purpose of the present study was to develop a questionary test which would have a practical value in

* This study was carried out at the University of Idaho. The writer wishes to express his appreciation to Professor J. W. Barton for his aid in the shaping of the study and to Professor D. G. Paterson of Minnesota for his help in preparing the manuscript.

differentiating individuals who are mentally unstable from those who are in good mental health. The plan for the development of such a test called for the inclusion of such questions as imply the presence of a severe maladjustment of personality; it further demanded a list of questions having a high degree of internal consistency; and finally, the plan necessitated an evaluation of the test against an outside criterion.

Previous Studies

Woodworth, in connection with his World War service, made the first attempt to detect mental instability by means of a group test; and Cady (3), House (7), and Papurt (9) have prepared revisions of the Woodworth test. Laird (8) developed a test of mental instability which is largely made up of the questions used by Woodworth but which differs markedly in the way these questions are to be answered and scored. Thurstone and Thurstone (10) recently described a similar test of mental instability. They included most of the questions of the Woodworth test but added others. At a still more recent date, Bernreuter (2) published a similar test. Because some of the tests have not been validated, only those which entail some index of validity will be discussed further.

Franz (4) in his manual for conducting mental examinations describes the Woodworth Psychoneurotic Inventory and in discussing norms states: "Normal individuals (white) answered wrongly about ten questions of the total 116. Probably any individual who answers 20 of the questions wrongly should be suspected of instability. * * * If the number of 'wrong' answers is greater than 30, grave suspicion of abnormality is warranted." But one would not be warranted in using Franz's norms since the factor of motivation was not taken into consideration. The importance of motivation was demonstrated by Hollingworth, (6) who compared the Woodworth test

scores of psychoneurotic recruits tested before the armistice was signed with the scores made by similar recruits after the armistice was signed. In respect to the pre-armistice group he says: "The most probable score (medians) of white recruits was 10 points, colored recruits 19 points. Cases of neuro-cardiac-asthenia, psychoneuroses, and other disorders encountered in the recruiting service averages from 30 to 40 points." But with respect to the post-armistice group he states: "The post-armistice group shows a distinctly different distribution. Nearly 50 per cent are below 10 points, the median score lying just above that score." The difference in the scores of the two groups of psychoneurotics was interpreted as being due to differences in motivation. If this conclusion is valid, it seems reasonable to expect the Woodworth test to yield variable results for normals and psychoneurotics when they are tested in test situations where strong motivation to reveal or to conceal mental and somatic difficulties is likely to be present.

The above conclusion is further supported by two later studies of the Woodworth test. Garrett and Schneck (5) in comparing psychoneurotics and presumably normal college students found that the psychoneurotics exhibited only a slightly greater frequency of atypical answers. Papurt (9) in studying 100 insane, equally divided as to sex, found the average score of the males to be 19.34 and of the females to be 19.92. These averages do not seem to be significantly high when it is noted that Garrett and Schneck report for Columbia College freshmen an average Woodworth test score of 16.67. Papurt has also indicated that for the insane only 14 per cent of the males and 10 per cent of the females equalled or exceeded a score of 30.

House (7) in publishing his revision of the Woodworth test claims that the test is valid since groups of college students average 25.09 atypical answers, whereas a group of avowed psychoneurotics averages 38.13. But

he fails to report the variability of scores made by each group. Furthermore, the central tendency for college students shifts from one college to another. For example, 102 students at West Point average 20.98 atypical answers as compared with an average of 30.31 given by students at the College of the City of New York. This difference between the average scores of these two groups, both of which were labelled by House as "normal," is almost as great as the difference between the average scores of the total group of "normals" and psychoneurotics. The data presented by House do not preclude the possibility that the overlapping of the scores for the "normals" and psychoneurotics was so great that a low degree of validity actually obtains for his test.

The questionary developed by Thurstone and Thurstone (10) included most of the questions of the Woodworth test and some additional ones to bring the total number of questions to 223. From test results of 694 college students, they have attempted to show that the test is valid by the following method: on the basis of total atypical answers these writers selected the 50 highest scores and the 50 lowest scores, designating the former as "most neurotic" and the latter as "least neurotic."

Each item was then statistically analyzed to determine the number and per cent in each group who had answered atypically. It was found that for each item the group classified as "most neurotic" gave a higher percentage of atypical answers than did the "least neurotic" group, thus demonstrating that the questionary is composed of a highly consistent set of questions. However logical it may seem from a statistical standpoint, their claim that the test is valid because it is internally consistent is only an assumption. Granting for the moment that mental instability has been measured to some degree by the test, there are no figures given which demonstrate that this degree of validity is high enough to

insure an acceptable discrimination between normals and abnormals. Furthermore, the available evidence indicates that the Woodworth test is of little value in making such discriminations. The Thurstones have made use of practically all the items employed by Woodworth and there is no reason to assume, without further evidence, that these items are of more value when used in the Thurstone test than in the original.

Bernreuter's personality test is also supposed to yield a measure of neurotic tendency. As evidence for validity he reports a perfect positive correlation (corrected for attenuation) between his measure and the Thurstone Neurotic Inventory. But such evidence is not adequate since we require some "outside criterion" against which to validate such questionnaires.

The above analysis of the previous work done with questionary tests of mental instability is intended to show that *as yet no such test has been adequately validated.* At the same time, the writer does not claim that the tests discussed here are lacking in validity, although some evidence points in that direction. The burden of proof rests on those who have constructed the tests.

METHODS

1. *Construction of the Test.*

The development of earlier forms of the present test has been described elsewhere: (I) Three criteria were employed in the selection of diagnostic items: (1) an atypical answer to each question must occur in less than 25 per cent of the total group tested; (2) the set of questions must be highly consistent; and (3) each question must show definite value in discriminating between groups of college students classified on non-test bases as "stable" and "unstable." The percentage of atypical answers to a question given by the "unstable" group must be five or more times greater than the percentage of atypical answers given by the "stable" group in order

for that question to be included in the scoring key. On the basis of these criteria, 84 test items were selected.

Two forms of the test were then devised. Form 1 was constructed as follows: The items which had been selected by the foregoing procedures were again put into questionary form. Each question was to be answered by *Yes* or *No*. In each case, the atypical answer was *Yes*. The test score equalled the number of atypical answers given. The following are sample questions from Form I: "Do you frequently faint away?" "Are you constantly unhappy?" "Have you ever had a severe nervous breakdown?" "Is sex the most hated and dreaded part of your life?" "If your subsistence (food and shelter) were provided for, would you prefer to daydream and live in your world of fancy rather than in the world of reality?" "Do you constantly observe and overhear other people making slighting remarks about you?" and "In your daydreams do you frequently imagine yourself to be some divine being as God or Christ?"

Form 2 was constructed as follows: The same test items employed in Form 1 were used and were arranged in the same order. In this case each question merely implied a "symptom" and made no reference to the frequency of its occurrence or its degree of severity. A means was provided whereby the subject was to rate himself as to the severity of each "symptom" and as to the frequency of its occurrence.

Space precludes a complete description of the elaborate scoring method used for Form 2 but it will be characterized briefly. The "symptoms" implied in the test were each assigned "basic weightings" of from 1 to 4. The degrees of the severity which could be indicated for each symptom were classified into four groups. The group including the degrees of least severity was weighted as 1, the second as 2, the third as 3, and the fourth, or highest, a weighting of 4. These weightings were referred to as "severity weightings." In a similar manner the fre-

quencies with which symptoms were reported to occur were grouped in a fourfold classification and received "frequency weightings" of from 1 to 4. The score for each item is equal to the product of the "basic weighting," the "severity weighting," and the "frequency weighting" multiplied together. The score for the test equalled the sum of the item scores.

The cover-page was identical for each of the test forms and included an explanation of the purpose for which the test was given. The test was described as an "inventory of personal problems." The explanation was further elaborated in a way that was intended to aid in putting the subject at ease and to encourage truthful answers. The directions for answering the questions were located at the top of the second page in each of the two test forms. The following is an excerpt from the directions on Form 1: "Do not give the answer *Yes* to a question unless you actually have the problem to the degree which the question implies. If you are in doubt about the answer to a question give the answer *No*. On the other hand do not avoid giving the answer *Yes* if you feel that it best describes you."

2. Administration of the Test.

Both Form 1 and Form 2 of the questionary were filled out by 130 General Psychology students at the University of Idaho. The students were instructed by the writer to take alternate seats in the lecture-room and upon receiving a copy of the test to read very carefully the explanation and directions given. Form 1 was filled out by the students and returned immediately upon completion. The procedure was then repeated for Form 2 during the same class period. The directions on Form 1 were followed satisfactorily and no questions concerning them were asked. It was necessary to answer several individual questions regarding the mechanics of answering the questions in Form 2.

3. Criterion of Test Validity.

The criterion of the validity of each of the two test forms was a personality rating of each individual derived from personal interviews and information provided by deans, instructors, and class-mates. A conference was held with each individual before the rating was made; each subject was asked to come to the experimental room and take a galvanometer test of "emotionality." A free association test was given the subject and the psychogalvanic deflections on a Hathaway galvanometer were recorded for each response. This test was not given for its own sake, but was used as a method of getting the subject into a situation whereby he might be easily led into discussing his personal life.

When the galvanometer test was completed, the subject was asked whether or not he objected to being asked a number of questions concerning his personal life. None of the subjects objected. The conference was usually begun with a discussion of "emotionality;" matters of irritability, restlessness, inability to sleep well, etc., were brought up. From this point the discussion led to the more serious symptoms of instability. When a symptom was uncovered it was discussed fully with the subject regarding its frequency, severity, origin, the attention given to it, etc. Every attempt was made to avoid placing a stigma on any problem acknowledged by the subject. The frankness with which each individual discussed his personal life indicated that this attempt was largely successful. The length of the conference varied from 30 minutes to 3 hours; the average length of time spent with each individual was approximately 60 minutes.

On the basis of the conference results, observation of the social behavior of the subject, and confidential reports by instructors, each individual was rated as belonging in one of five groupings according to the "degree of instability" thought to be present. At the time these

ratings were made, the Mental Instability Tests previously taken by the individuals rated had not been scored. A definition of each group together with illustrative cases follows:

a. *Individuals who give every evidence of being well adjusted.*—W. S. Male student. Age 19. Physical condition excellent. Has a reputation for coolness and steadiness on the football field. Well adjusted in his social life. Is friendly. Popular with others. Always appears at ease. Discussed his personal life quite frankly. Gives evidence of being relatively free from tendencies which might be indicative of instability.

b. *Individuals having only mild maladjustments.*—H. N. Male student. Age 20. Physical condition excellent. Somewhat introverted. Rather quiet. Was somewhat nervous during the conference. Apparently quite willing to discuss himself in respect to the questions asked. Said that he is easily embarrassed. Likes to be alone for part of the time. Has a mild dread of meeting people. In general he enjoys the company of others and is apparently well adjusted socially.

c. *Individuals having personality maladjustments which definitely handicap them in some respect.*—H. D. Male student. Age 20. Physical condition excellent. High school record excellent. His scholastic record is above average but he is discouraged with himself for not having done better. Finds no subject which really interests him. Is unable to concentrate effectively. Is easily depressed. Feels that he is doing wrong in accepting money from his relatives when he will probably not be able to apply his college education successfully. Worries about religion. Is mildly nervous. Has a strong fear of life and lacks self-confidence in matters of working toward his ideals. Has made a good social adjustment and has the necessary self-confidence in social situations.

d. *Individuals having well developed tendencies to mental instability.*—T. C. Female student. Age 23. Physical health fair. Nervousness manifest. Excellent scholastic record. Not well adjusted socially. Social qualities rather poor. Has few real friends. Frequently has severe dizzy spells. Severe shaking spells occur when she is in an unusual situation. Experiences short periods of unconsciousness occasionally but does not fall. Several times per day she experiences unpleasant taste sensations which are quite vivid. Day-dreams a great deal and prefers this to reality. Complains that she has been subjected to a great deal

of ill treatment by others. Frequently hears others making slighting remarks about her. These beliefs have some basis in fact. Experiences pseudo-hallucinations. Can see persons who are either relatives or friends quite vividly whenever she attempts to do so. These visions have never appeared except through her conscious effort to produce them. Although the visions are as vivid as in reality, she never interprets them as being real. Sometimes feels that thoughts and ideas are being forced into her mind by another person. Believes that this takes place through mental telepathy. Such thoughts occur infrequently and she does not worry about them.

e. *Individuals who have at the present time rather seriously established syndromes of mental disease.*—B. K. Female student. Age 22. Physical condition poor. History of nervous breakdown while in high school. Breakdown was severe and recovery has been incomplete. Insomnia present. Constantly feels tense, uneasy, and unhappy. Frequently has recurring unpleasant dreams. Each concerns an accident. Feels that she does not make friends with others but is not concerned about this as she tires of people readily. Hates most people. Trusts no one. Frequently hears voices which she cannot distinguish from reality until she investigates the source of the voice. The voices command her to perform senseless acts. Sometimes hears sounds such as chimes or music which she cannot distinguish from reality. Frequently has vivid sensations of something crawling on her body. Knows that it is a "little black bug" although she has never been able to find it. Believes that someone has a strange power over her and forces strange thoughts into her mind. Cannot identify the person. Dreads life in general and has planned to commit suicide. Is afraid to go through with the plan. Notwithstanding the seriousness of her symptoms she gives every appearance of being well adjusted in her social life. There are few manifest symptoms of instability in her overt behavior. She stated that she had never spoken to anyone concerning herself in this respect and that she had made every attempt to give the impression that she was happy and well satisfied with life.

III. Results

The test papers of the 75 individuals who were interviewed and rated were then scored. Since six students had obviously made no attempt whatever to answer the questions in Form 2 of the test, the results for Form 2 are based on 69 cases.

Table 1 indicates the extent to which the five classes of students are differentiated on the basis of mean score in the two forms of the Mental Instability Test.

TABLE 1

Class Means for Form 1 and Form 2

	Form 1		Form 2	
Personality Ratings	Mean Score	No. of Cases	Mean Score	Number
Class a.	0.00	2	0.00	2
Class b.	0.70	27	8.25	27
Class c.	2.24	29	22.07	29
Class d.	7.85	14	77.74	8
Class e.	19.33	3	233.33	3

Mere inspection of Table 1 indicates that each form of the test discriminates fairly well among the criterion groups. Since the test was made up of questions which concerned only the more serious "symptoms," it was expected that a large number of the test scores would be zero. This explains the lack of test discrimination between classes a and b.

Classes a, b, and c, will now be considered as one group and will be termed "stable," whereas classes d and e will be considered as "unstable." Tables 2 and 3 demonstrate how well Form 1 and Form 2 respectively discriminate between the two groups.

TABLE 2

Means and Differences between the Means on Form 1

Group	Number	Mean Score	PE_m	Diff.	$PE_{diff.}$	$\frac{D}{PE_{diff.}}$
"Stable"	58	1.466	.156			
"Unstable"	17	9.765	.900	8.299	.913	9.09

TABLE 3

Means and Differences between the Means on Form 2

Group	Number	Mean Score	PE_m	Diff.	$PE_{diff.}$	$\frac{D}{PE_{diff.}}$
"Stable"	58	20.34	1.53			
"Unstable"	11	126.82	15.09	106.48	15.16	7.02

The above results show that both Form 1 and Form 2 yield a statistically reliable difference between the "stable" and "unstable" groups.

The correlation between the scores on Form 1 and a four-fold classification of the subjects was computed by means of the Contingency Coefficient. Since the test was not intended to discriminate between Class a and Class b, and since little would be gained from such discrimination, these two groups were treated as one class in the statistical procedure. The Contingency Coefficient, computed by the short method was found to be .76. The Contingency Coefficient between the same four-fold grouping of the subjects and scores on Form 2 was also computed and found to be .81. When the Contingency method is used with as small a number as four categories, as in this instance, C cannot be considered as the equivalent of r but only as a rough approximation to it. Furthermore, the difference between C .76 and C .81 as determined by this procedure cannot be considered to be statistically reliable because of the large probable error involved.

A Pearson r between the results of Form 1 and Form 2 was found to be $+.87\pm.01$. This correlation is high enough to insure a satisfactory reliability for each form of the test.

IV. Discussion

The studies of Garrett and Schneck (5) and of House (7) indicate that only the severe maladjustments are indicative of mental instability. In Form 1 of the present test the questions concern "symptoms" which the subject believes to be present to an extreme degree and to be of frequent occurrence. In Form 2, the subject is asked to indicate the frequency with which each symptom occurs and also to indicate its severity. The answers were scored in such a way as to give a much higher weighting to "symptoms" of frequent occurrence and of severe form than to those of infrequent occurrence and mild form. Form 1 and Form 2 of the Mental Instability Test, differ essentially from all other tests of "instability" in two ways: 1. Most of the "symptoms" implied by the questions are different from those implied by the questions used in other tests; 2. The magnitude of the test scores is determined largely by the acknowledgment of "symptoms" which occur frequently and in severe form. It is reasonable to assume that the items employed in other questionary tests of this nature concern maladjustments which are of too mild a form and of too frequent an occurrence in relatively stable individuals to be of much diagnostic value in identifying the unstable.

Each of the two test forms has been shown to be reliable and to exhibit a fairly high degree of validity. Form 1 is easier to answer and to score than Form 2 and is, therefore, to be preferred. There appears to be little difference in the validity of the two forms.

A serious criticism may be directed toward the criterion utilized in validating the test, namely, incapacity on the part of the writer to make correct diagnoses of mental instability on the basis of interviews and reports of instructors. It must be admitted that the writer's previous experience in a diagnosis of mental disease and training in abnormal psychology was by no means equiv-

alent to that required of a psychiatrist. Furthermore, a conference of 60 minutes is not adequate for a complete mental examination. Although the test scores made by each subject was unknown at the time the "diagnosis" was made, it is possible that the "diagnosis" itself was based largely on "symptoms" revealed in the interview which were identical with those tapped by the test questions. Some neurotic subjects may have successfully concealed their "symptoms" in both the test and the conference. For these reasons it is possible that the diagnoses may have only a low degree of validity.

Without claiming that the present Mental Instability Test is a superior instrument for "spotting" mental instability, the evidence cited in its behalf represents a sincere effort at validation. Obviously, additional experiments should be made in which psychiatrists are employed in making the necessary diagnoses among "normal" subjects for validation purposes or in which the responses of avowed psychoneurotics are compared with the responses of normals.

V. Summary

A critical review of the literature has led to the conclusion that satisfactory evidence as to the validity of questionary tests of mental instability is lacking. An attempt to construct and validate such a test is described. The test items were selected on the basis of frequency of occurrence in groups of "normal" college students and value in supporting the internal consistency of trial sets of questions. The hypothesis for test item selections was that a "symptom" must be of frequent occurrence and of considerable degree of severity to be of value in discriminating between stable and unstable individuals. Eighty-four test items were selected and two forms of the questionary test made out, each employing the same items but differing in respect to the method used in answering the questions. The degree of correlation be-

tween the scores made by 69 students on each form of the test was r + .87±.01 and is high enough to insure a satisfactory reliability for each form. Validity was measured by calculating the Contingency Coefficients between the scores on each form of the test and a four-fold classification of subjects on the basis of the degree of mental instability judged by us to be present. C was found to be +.78 and +.81 for Form 1 and Form 2 respectively, thus demonstrating a fairly high degree of validity as determined by this criterion. Our inability to meet possible criticisms of the criterion we employed necessitates further evaluations of the test, preferably in terms of agreement with results provided by psychiatric examinations of subjects tested.

BIBLIOGRAPHY

1. Barton, J. W. & Ingle, D. J. "Detecting Mental Instability." *Jour. Ab. & Soc. Psych.*, 1932, 27, 52.

2. Bernreuter, R. G. "Manual for the Personality Inventory." *Stanford University Press*, Stanford University, California.

3. Cady, V. M. "The Estimation of Juvenile Incorrigibility." *J. of Delinquency*, Monog. No. 2, April, 1923.

4. Franz, S. I. *Handbook of Mental Examination Methods.* New York, Macmillan, 1919.

5. Garrett, H. E. & Schneck, M. R. "A Study of the Discriminative Value of the Woodworth Personal Data Sheet." *J. of Gen. Psych.*, 1928, 1, 459.

6. Hollingworth, H. L. *The Psychology of the Functional Neuroses.* New York, D. Appleton & Co., 1920.

7. House, S. D. "A Mental Hygiene Inventory." *Archives of Psych.*, 1927, 14.

8. Laird, D. A. "A Mental Hygiene and Vocational Test." *J. of Ed. Psych.*, 1925, 16, 419.

9. Papurt, M. J. "A Study of the Woodworth Psychoneurotic Inventory with Suggested Revision." *J. of Ab. & Soc. Psych.*, 1930, 25, 335.

10. Thurstone, L. L. & Thurstone, T. G. "A Neurotic Inventory." *J. of Soc. Psych.*, 1930, 1, 3.

THE RELATION BETWEEN PITCH DISCRIMINATION AND ACCENT IN MODERN LANGUAGES

EMILY S. DEXTER and KATHARINE T. OMWAKE
Agnes Scott College

The writers have speculated for some time as to a possible relation between ability to discriminate pitch and ability to pronounce well in French, and perhaps in other foreign languages. This study was undertaken with a view to getting at least a partial answer to the question.

Two of the Seashore Measures of Musical Talent, those for Pitch Discrimination and for Tonal Memory, were given to an unselected group of 118 upper class students at Agnes Scott College, all but two of whom had had French in college. (Practically all had also had two years of French in high school.) About 25 had had German also and about 20 had had Spanish. Ratings as to pronunciation and accent of each individual tested were obtained from members of the French, German, and Spanish departments. These ratings were in five gradations; 1 being high, 3 average, and 5 low, with 2 and 4 as intermediary steps. There were 95 students for whom French accent ratings were secured, rated by three members of the department, but only 20-25 in Spanish and German; so the study has concerned itself mainly with French.

Since no apparent contribution toward accent was made by the test for Tonal Memory, its results were discarded and will not be discussed further in this paper.

In order to see whether certain other aspects of the individual, such as intelligence and personality, entered in

to affect pronunciation, measures of those factors were obtained. To test intelligence Form B of the Henmon-Nelson Tests of Mental Ability for College Students was administered. To investigate personality Bernreuter's Personality Inventory was given. Since, according to Bernreuter the scale for Neurotic Tendency (B1-N) correlates about .93 with the Introvert-Extrovert scale (B3-I) we did not take scores from it into account, but used only the scales for Self-Sufficiency (B2-S), Introvert - Extrovert (B3-I) and Dominance - Submission (B4-D). All results from the intelligence and personality tests are reported in percentiles.

Another question raised was whether there is any relation between the number of years of French studied and pitch discrimination.

The findings of the study were as follows. First, in regard to the question last raised, it was found that none of the 29 in the lowest quartile in pitch discrimination and only 28 per cent of those in the next to the lowest quartile had taken more than two years of French in college. In other words, only 12 per cent of the students, that is, 8 out of 58, below the group median in pitch discrimination had taken more than two years. The two students who had had no French even in high school chanced to be the lowest in the group. In the following table are to be found the median scores for the groups together with the percentile ranks assigned these scores by the Seashore standards which tend to magnify the differences between the scores.

However, there were nearly as many students with good pitch discrimination who took two years or less as there were students with poor pitch discrimination. Evidently a student may have good ability in pitch discrimination and not take much French. But he may not have poor discriminative ability for pitch and study many courses in French. The correlation between years of French and pitch discrimination was $.225 \pm .059$.

Median Scores and Percentiles on Pitch Discrimination for Varying Amounts of French

	Number of Years of French			Entire
	3-4	0-2	0-1	Group
Median score	81	76	73	78
Percentile	50	29	21	36
No. cases	26	90	43	116

As to the relation between rating in accent or pronunciation and ability to discriminate in pitch, only 5 out of the 48 who were below the median in pitch discrimination were rated above the average in French accent, and of those five only one, rated 2, was in the lowest quartile; whereas 15 of the sub-median group were rated lower than average. The other 28 were average. Therefore, it appears that a student does not have low ability in pitch discrimination and at the same time a good French accent. However, many of those making high scores on the pitch test were rated low in accent. To offset the group of 15 just mentioned who were below the median in pitch discrimination and rated low in accent, there were 17 above the median who were rated low. The median percentile on the pitch test of those rated high in accent was 56, of those rated average, 36, and of those rated 4, 50; but of those rated 5 the median percentile in pitch discrimination was 23. The correlation between accent rating and pitch discrimination was .210 $\pm$ .066.

In order to see what effect intelligence had, the correlation between pitch and intelligence and between intelligence and accent ratings were calculated. The former coefficient was —.103 $\pm$ .085, and the latter was .485 $\pm$ .075. Intelligence very evidently affects the accent rating but not the ability to discriminate pitch. The relation between pitch and accent with intelligence eliminated is indicated by a partial correlation of .300, showing that intelligence plays only a small part.

In German and in Spanish the numbers rated were so small that the groups rated 2 and 3 were the only ones

large enough to consider. In Spanish the group rated 2 (only 8 in number) had a median pitch discrimination of 81, i.e., percentile 45, whereas the group rated 3 (only 9 in number) showed a median of 70, i.e., percentile 15. The two groups made the same average scores on the intelligence test. While this is entirely too small a number from which to draw a reliable conclusion, it appears that given enough mental ability to do college work in Spanish, ability to discriminate pitch contributes more toward a good accent than does intelligence.

In German, 9 individuals were rated 2 in pronunciation, and 14 were rated 3. The two groups had the same pitch median, but the median intelligence percentile of the higher group was 20 points above that of the lower. Hence, while again the numbers are too small for generalization, in German intelligence seems to contribute more toward accent than does ability to discriminate pitch.

Turning next to the relation between the personality tests and French pronunciation, no relation was found between accent and scores on the test for dominance. A slight tendency was apparent for students making low scores in self-sufficiency to be rated low in accent, and for those making high scores to be rated high. Out of 10 individuals rated better than average in accent, 3 made low scores and 7 high in self-sufficiency. In fact the 6 students who were above the 70th percentile in self-sufficiency were all rated average or above; while of those scoring below the 70th percentile the ratings were scattered.

The test for introversion-extroversion yielded much the same type of relation. Those rated high in accent tended to score high on introversion, 7 of the 10 rated above the average being above the 70th percentile. However, nearly as many who were high in introversion were rated low in accent as were rated high, but only 2 students low in introversion were rated high in accent.

SUMMARY

The study seems to show:

1. A fair relation between intelligence and accent ratings in French. ($r = .485 \pm .075$)

2. A slight relation between pitch and number of years of French. ($r = .225 \pm .059$)

3. A slight relation between accent and pitch discrimination. ($r = .210 \pm .066$)

4. No relation between intelligence and pitch discrimination. ($r = -.103 \pm .085$)

5. No relation between tonal memory and French accent.

6. An appreciably higher median in pitch discrimination for the group rated above the average in accent in French and Spanish than for those rated average or lower.

7. An appreciably higher median in pitch discrimination for the group having had 3 or 4 years of French than for those having had 2 years or less.

8. Intelligence to be a prominent contributing factor in the securing of good accent rating in French and German.

9. A slight tendency for those low in introversion not to be rated high in accent, and for those high in introversion to be rated high in accent oftener than low.

General Conclusions: People with high ability in pitch discrimination may be rated either high or low in French accent and take much or little French in college, but people with low ability in pitch discrimination are not rated high in accent, nor do they take more than two years of college French.

A MORE VALID METHOD OF COMPARING THE ACCOMPLISHMENT QUOTIENTS OF INDIVIDUAL PUPILS WITH THOSE OF THE GROUP

ROBERT A. DAVIS and WILLIAM A. CAMPBELL
University of Colorado

Progress in education is dependent upon the solution of new problems and the refinement of methods used in their solution. In the effort to make contributions to the content of education there has been a decided lack of systematic effort to establish and refine the methods, tools, and devices which are used in the solution of educational problems. The findings of investigations, however, are valuable only insofar as they are based upon reliable tools and devices used in obtaining and interpreting data.

The value of the accomplishment quotient as a tool for teachers and research workers has been the subject of much controversy. Several attempts have been made to justify its use and to defend it as a valid educational concept. The value of the accomplishment quotient will depend upon the degree to which improvements overcome inherent defects of the quotient technique. While earlier investigations dealt with the advantages of the accomplishment quotient, the later studies have tended to study critically the AQ technique and have warned teachers against too great dependence upon its validity as an index between intelligence and achievement.

Toops and Symonds[1] refer to the AQ as a measure for combining effectively the results of educational and

[1] Toops, H. A., and Symonds, P. M., "What Shall We Expect of the AQ?" *Jr. Ed. Psy.*, 13, 513-5, 1922, 14, 27-28, 1923.

mental tests into a measure of achievement relative to the pupil's capacity to learn. Stebbens[2] believes that the AQ is the fairest and most valuable measure of the efficiency of both teacher and pupil, because it evokes from pupils and teachers results which are in proportion to the native ability of the pupils. McPhail[3] notes that bright pupils are less apt to work in accordance with capacity than the dull, and that the dull child achieves more than is expected when capacity is taken into account. Several investigators have noted the negative correlation which usually exists between IQ and AQ and have suggested various explanations to account for it. Wilson[4] believes that the negative correlation is due to spurious factors entering into the computation of quotients as well as the unreliability of tests upon which the AQ is based, and that intelligence is not the sole determinant of achievement. Torgerson[5] and Shuman suggest that the negative correlation may be partially due to improper classification. Goodenough[6] states that learning is a function of time as well as ability and that a child with a mental age above the norm for his grade does not have opportunity to acquire enough information to raise his AQ to the desired level, which implies that most classes are taught for the average pupil. These points of view do not exhaust the literature on the subject but are presented to give a general review of the advan-

[2] Stebbens, R. C., "Accomplishment Quotients as an Aid in Diagnosis," *El. Sch. Jr.*, 8, 382-92, 1929.

[3] McPhail, A. H., "Correlation Between IQ and AQ," *School and Soc.*, 16, 1922, 586-8.

[4] Wilson, W. R., "The Misleading AQ," *Jr. Ed. Res.*, 17, 1928, 1-10.

[5] Torgerson, T. L., and Shuman, Irene, "Variability of Accomplishment of Pupils of the Same Mental Age," *Jr. Ed. Res.*, 11, 1925, 132-36.

[6] Goodenough, F. L., "Efficiency of Learning and the Accomplishment Ratio," *Jr. Ed. Res.*, 12, 1925, 297-300.

tages and limitations of the accomplishment quotient as revealed by investigators.

The accomplishment quotient has gained recognition chiefly on the ground that it is an effective measure of the relation between ability and achievement. If the AQ fulfills the requirements of a desirable instrument, effort should be made to eliminate its objectionable features. Some objections may be overcome by refining the measures of intelligence and achievement; others may be eliminated by training teachers in the proper use of standardized tests. In addition, two other obstacles which are inherent in the technique should be corrected. One disadvantage is the spurious element entering into the computation of quotients and the other is the fact that the AQ must take into account mental age and rate of mental growth in relation to level of instruction. This paper does not obviate these difficulties but proposes a more valid means of comparing the achievement quotients of individual pupils with those of the group, and should make the AQ a better expression of the relation between intelligence and achievement.

Sources of Data and Procedure

Since the usual technique yields an expected accomplishment quotient too high for bright pupils and too low for those who are dull, the proposed method operates to make the expected AQ dependent upon the capacity and achievement of the group rather than of some theoretical norm. Thus bright pupils will be expected to have accomplishment quotients of less than 100 while dull pupils will be expected to have those of more than 100. The expected AQ may be computed from a regression line of IQ on AQ. The IQ's and AQ's are determined in the usual manner. The index of achievement may then be found by subtracting the actual AQ from the expected AQ. Obviously by this procedure the expected

AQ's for different groups will vary, the amount of variation depending upon average ability and achievement.

In order to study the actual AQ expected, AQ procedure tests were administered in the sixth grade of fifteen schools in northern Pennsylvania. To determine IQ's the Haggerty Intelligence Examination, Delta 2, was used; and to determine EQ's the New Stanford Achievement Test, Form V, was employed.

The first step in the treatment of the data was the determination of various measures of central tendency and variability. This information should show how the different classes compare as to general ability and achievement. It will be noted from Table I that the mean IQ for the entire fifteen classes is 98.89 which closely approximates the theoretical average of 100. The range in mean IQ for the schools is from 89.41 to 107.88, showing a variation of 18.47 points. From the same table, the mean AQ is found to be 100.18 and the range for the separate schools, from 95.38 to 106.32, a difference of 10.94. It is interesting to note that the school with the lowest average IQ (89.41) has the highest average AQ (106.32) while the group with the highest average IQ (107.88) has the lowest AQ (95.38).

The schools used in this study show marked variation both in intelligence and achievement. The mean IQ's range from schools of superior to those of inferior ability. Standard deviations for both IQ and AQ also show much variation indicating that the schools represent a fair sampling of abilities. This condition is to be expected as there is little selection until after the sixth grade.

When we turn to the relation between intelligence and achievement quotients it may be observed that in all cases except one negative coefficients are found. These results agree with those of other investigators who have studied the relationship of these measures.

Omitting the exception the negative coefficients range from –.313 to –.831. The coefficient for the entire number of cases is –.511.

In order to study some of the factors affecting these results, coefficients were computed to determine the relationship between chronological age and accomplishment quotient and between intelligence quotient and chronological age. In the first case a coefficient of .156 resulted showing that there is little, if any, relation between the two factors. In the second case, a coefficient of –.627 was found. A fairly significant relationship would be expected here as chronological age is a factor on which the IQ is based and IQ correlated with AQ to the extent of –.511.

Since the chronological age for the group varied from 116 months to 198 months, the next step was to determine whether wide variability in ages affected the rela-

TABLE I

Results of Treatment of Scores to Obtain Mean IQ, AQ, σ_{IQ}, σ_{AQ}, and Correlation Between IQ and AG

School	Number of Pupils	Mean IQ	Mean AQ	σ_{IQ}	σ_{AQ}	r	PE
1	36	104.02	100.8	18.9	7.5	—.707	±.057
2	36	99.58	96.39	16.2	8.55	—.504	±.085
3	34	102.94	102.9	16.4	8.65	—.685	±.062
4	34	89.41	106.32	13.75	7.25	—.507	±.086
5	33	93.86	99.46	16.5	11.2	—.497	±.088
6	33	90.38	103.86	12.95	6.85	—.591	±.075
7	31	98.95	98.11	17.6	8.5	—.732	±.053
8	30	95.66	99.5	18.6	8.6	—.409	±.103
9	28	102.5	100.	18.25	7.8	—.313	±.117
10	26	105.58	97.69	21.3	7.4	—.497	±.100
11	26	107.88	95.38	19.25	6.2	—.419	±.112
12	15	92.17	101.17	18.55	6.4	—.747	±.079
13	12	98.33	101.25	14.65	5.05	—.098	±.209
14	12	100.4	97.5	16.35	9.35	—.831	±.058
15	9	100.97	96.94	20.05	8.3	—.81	±.078
Total	395	98.89	100.18	18.15	8.8	—.511	±.025

tion between IQ and AQ. By means of partial correlations, the chronological age was held constant for the group. Although the correlation between IQ and AQ was –.511 in the first case, when age was held constant it became –.546.

In considering the significance of this change, the first observation is the extremely small difference in the two coefficients, indicating that the chronological age has little effect upon the relationship between IQ and AQ, and hence achievement. Instead of lowering the relationship it had a slight tendency to raise it indicating that if age is held constant for the group there is a tendency to reenforce the fact that achievement varies inversely with intelligence.

The Proposed Method

Since the accomplishment-quotient technique operates to raise spuriously the AQ's of those below average intelligence and lowers those above average intelligence but does not affect perceptibly those approximating the average, this method applies particularly to those cases of extreme ability who are most likely to be affected by any defect in the quotient technique. The actual AQ's are those found in the usual manner, while the term "expected AQ" as used here means those values found by a regression of IQ on AQ. This regression line is found from the correlation between IQ and the actual AQ and is plotted on the scatter diagram from which the correlation was originally computed. According to the usual AQ procedure, the theoretical ideal value for any IQ would be 100. By means of the regression line, the expected value for any IQ will vary and will not be a fixed norm. This value will vary according to the mean IQ and mean actual AQ of the group and will also take into account the correlation between the IQ and AQ and the variability of the two measures. In other words the expected AQ for any given IQ will

be based upon the measures of IQ and AQ and not upon some arbitrary value. Thus the expected AQ for any pupil is found in reference to the ability and accomplishment of the group. It will be in terms of what is expected of him under conditions of group ability and achievement.

The regression line of IQ on AQ for the group under consideration yields the following approximate values for the different levels of intelligence:

TABLE II

The Expected AQ's for Various Levels of Intelligence

IQ	Expected AQ
160	84
150	87
140	89
130	91
120	94
110	96
100	99
90	101
80	104
70	106
60	109
50	111

The middle fifty per cent of cases lie between IQ's of 85 and 110. The expected AQ's for this group range from 96 to 103. These facts show that the expected AQ for the group of average intelligence will vary but three points on both sides of the theoretical value of 100. The upper 25 per cent in ability range from 110-160. The corresponding variation in the expected AQ is from 84-96 showing a range of minus 12 points under the theoretical value of 100. The lower 25 per cent range from 51-85 in ability, and vary in expected AQ from 102-109 or plus 7 points. These results show that the correction influences the middle 50 per cent

by but 3 points on both sides of the ideal value of 100, for the upper 25 per cent it is affected by minus 12 points and for the lower 25 per cent by plus 9 points. The correction thus influences markedly the expected value of the AQ for pupils at the extremes of intelligence.

Table III shows the actual and expected AQ's for those pupils of superior and inferior intelligence. It will be noticed that according to the actual AQ technique that the upper 25 per cent failed to obtain the theoretical

TABLE III

Comparison of the Mean IQ, Actual AQ, Expected AQ and Differences for Upper and Lower 25 Per Cent of Cases

	Upper 25 Per Cent	Lower 25 Per Cent
Mean IQ	122.6	77.01
Mean actual AQ	94.3	105.46
Mean expected AQ	93.32	104.53
Difference in means of actual and expected AQ	.98	.93
Differences between the means of actual AQ and 100	—5.7	5.46

value by a difference of minus 5.7 while the lower 25 per cent exceeded it by a difference of plus 5.46. Under the expected AQ technique, the difference between the actual AQ and the expected AQ for the upper 25 per cent is .98 and for the lower 25 per cent .93. These facts show that the correction to the accomplishment quotient technique lowers the expectancy for the upper group, raises it for the lower group, and thus aids in making individual diagnosis of pupil ability more in accordance with the group than with a theoretical norm. The actual AQ expected AQ technique as a means of individual diagnosis within a group corrects the disadvantages of the usual method and makes individual achievement comparable (in a given group or class) although the individuals themselves may vary widely

in intelligence. However, individuals in one group are not comparable to individuals in another group as in both cases the expected AQ's are based upon different values. The technique is valid only when comparing individuals within a group.

The results of this investigation indicate that the actual AQ-expected AQ procedure overcomes certain limitations of the usual AQ technique. The corrected technique on the face of results obtained in this study operates to afford:

1. A more reliable means of comparing the achievement of individuals within a group, when the group as a whole varies widely in intelligence than was possible under the usual AQ method.

2. It also affords a more reliable method of comparing the achievement of groups when they vary widely in intelligence than was possible under the usual AQ technique.

Selected References

1. Abelson, Harold—"Achievement of Dull Pupils Under the Dalton Plan," *Sch. and Soc.*, 30, 1929, 211-12.

2. Coy, G. L.—"Study of Various Factors that Influence the Use of the AQ as a Means of Measuring Teaching Efficiency," *Jr. Ed. Res.*, 21, 1930, 29-42.

3. Franzen, Raymond—"The Accomplishment Ratio," *Columbia Univ. Contribution to Education* No. 125.

4. Goodenough, F. L.—"Efficiency in Learning and the Accomplishment Ratio, *Jr. Ed. Res.*, 12, 1925, 297-300.

5. Keys, Noel—"Shall Teachers Compute Quotients?" *Ed. Adm. and Supervision*, 17, 1931.

6. McPhail, A. H.—"Correlation Between IQ and AQ," *Sch. and Soc.*, 16, 1922, 286-8.

7. Popenoe, H.—"Report of Certain Significant Deficiencies of the Accomplishment Quotient," *Jr. Ed. Res.*, 16, 1927, 40-7.

8. Ruch, G. M.—"The Accomplishment Quotient Technique," *Jr. Ed. Psy.*, 14, 1923, 334-43.

9. Shriber, Joseph, and Davis, Robert A.—"The Ability and Achievement of the Pupils of El Paso County, Colorado," *Univ. of Colo. Bulletin*, 1929.

10. Stebbens, R. C.—"The Accomplishment Quotient as an Aid in Diagnosis," *El. Sch. Jr.*, 8, 1929, 382-92.

11. Toops, H. A., and Symonds, P. M.—"What Shall We Expect of the AQ," *Jr. Ed. Psy.*, 13, 1922, 513-5; 1923, 14-28.

12. Torgerson, T. L., and Shuman, Irene—"Variability of Accomplishment of Pupils of Same Mental Age," *Jr. Ed. Res.*, 11, 1925, 132-36.

13. Wilson, W. R.—"The Misleading AQ," *Jr. Ed. Res.*, 17, 1928, 1-10.

THE EFFECT OF SUMMER VACATION ON ABILITY IN TYPEWRITING*

LOWELL J. SCHROEDER
East High School, Sioux City, Iowa

This paper reports an attempt to find out the extent to which high school pupils lose their ability to manipulate the typewriter after a three months' summer vacation, during which time they do not use the machine.

The writer became interested in the problem when he heard a high school principal advise a pupil not to start typewriting at the beginning of the second semester as "All that is learned will be forgotten during the summer vacation period." It so happened that the writer was given a beginning class in typewriting of thirty-one pupils that same second semester. He, therefore, determined to find out, if possible, the effect that the summer vacation would have upon the speed and accuracy of these pupils.

Of the thirty-one pupils who registered for the class beginning the second semester, twenty returned the first semester of the following year to take the second half year's work under the direction of the writer.

Procedure During the First Unit of Work

Each pupil was allowed ninety minutes per day for typewriting. The entire time for the first six weeks was spent in teaching the class the keyboard and practicing on simple words and sentences. Of course no standard tests could be given until the entire keyboard had been taught. This was accomplished by the close

* Submitted by D. A. Worcester, University of Nebraska.

of the first six weeks. The first test, then, was given at the beginning of the second six weeks period. The plan was to give two tests each week, one on Tuesday and one on Thursday, and this was adhered to with very few exceptions. Ten tests were given during the second six weeks and eleven the third six weeks. The first test was for one minute, the second for two minutes, the third for three minutes, and so on up to the tenth, which was for ten minutes. Each of the remaining tests was for fifteen minutes. The tests were always given early in the class period, following a five to ten minute "warming up" exercise.

The Tests

The tests used in this experiment were the standard typewriting tests sent out by the various typewriter companies. No preference was given to the tests of any one company. They were chosen at random, and several of each were used during the semester. The tests were all graded according to International Typewriting Rules. Each pupil checked his paper at the close of the test. Each was then rechecked by the writer.

Procedure for the Second Half Unit of Work

The fall term began on Monday. The first test was given on Thursday. The intervening time was spent on routine matters such as checking out books, assigning machines, adjusting schedules, making out a laboratory budget, and so on. Some of the pupils "tried out" the different machines in an effort to find one that they felt suited them, but there was no scheduled or systematic practice of any kind. Each pupil was asked if he had used a typewriter during the vacation period and each answered in the negative. After this first test on Thursday tests were given regularly as in the preceding semester.

Data for Charts

The speed score, or net words per minute, as shown on Chart I was found by totaling the net number of words written by each pupil and dividing by the number who took the test. When a pupil was absent he was given a make-up test as soon as possible, either after school hours or during a vacant period.

The accuracy score, in terms of number of errors per minute, as shown on Chart II, was found by totaling

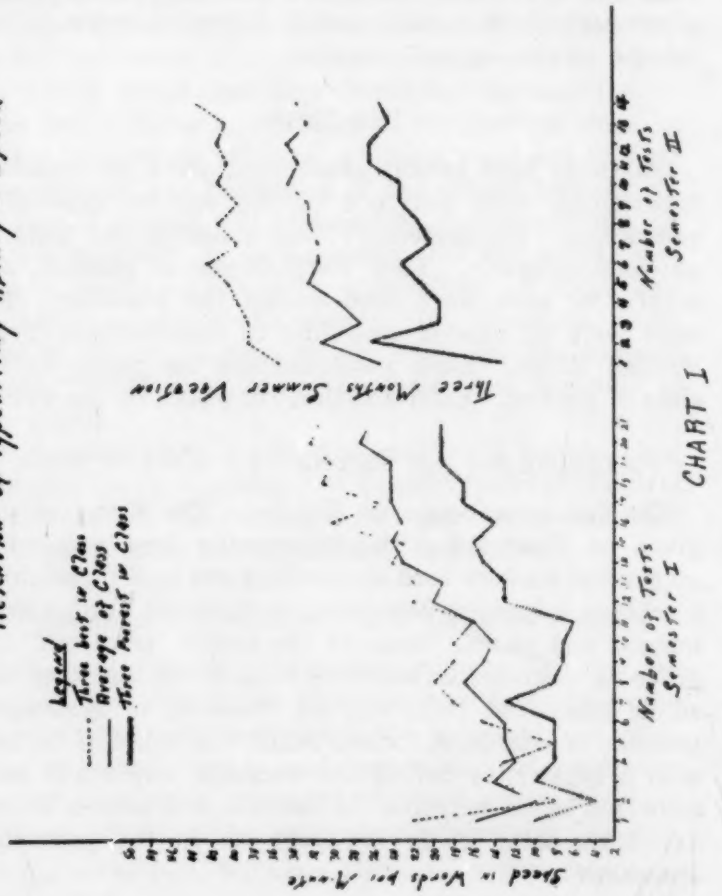


CHART I

the number of errors, dividing by the number who took the test, and then dividing by the number of minutes in the test.

Table I and Table II show individual gains or losses taken from the raw data.

THE RESULTS

Inspection of Chart I seems to show that the summer vacation period had very little effect upon the speed in

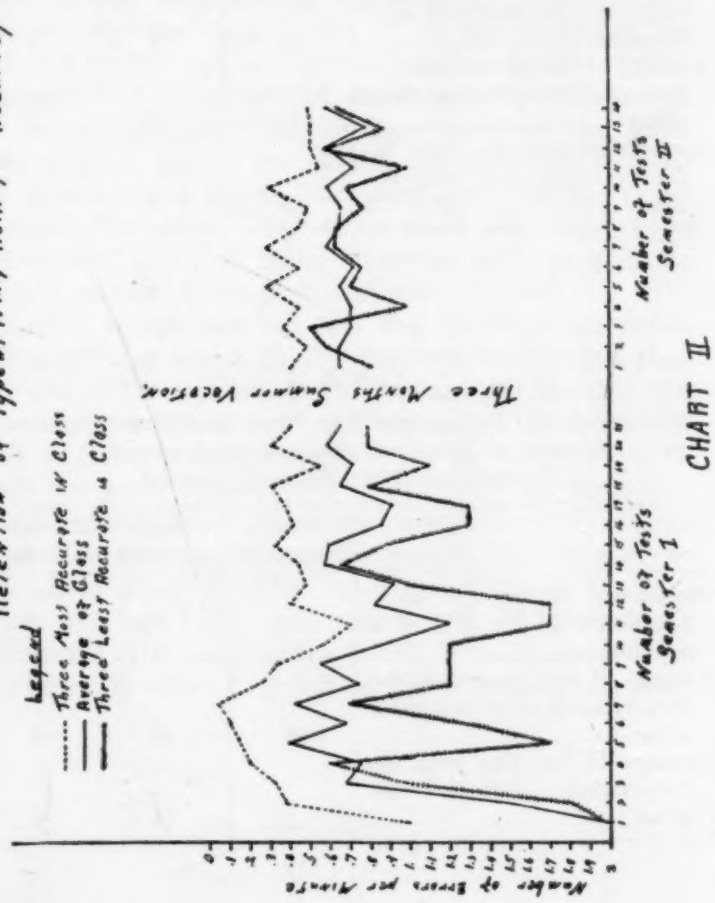


CHART II

TABLE I

COMPARISON

The Number of Individuals Who Increased, Decreased, or Remained Constant in Speed of Typewriting After Summer Vacation

	Number Increased Speed	Number Decreased Speed	Number Remained Same
Comparison of last test before vacation with first test after vacation	12	7	1
Average of last three tests before vacation with first three after	16	3	1
Average of last five tests before vacation with first five after	19	1	0

TABLE II

The Number of Individuals Who Were More Accurate, Less Accurate or Remained the Same in Accuracy in Typewriting After Summer Vacation

	More Accurate	Less Accurate	Remained Same
Comparison of last test before vacation with first test after vacation	11	9	0
Average of last three tests before vacation with first three after	10	10	0
Average of last five tests before vacation with first five after	9	11	0

typewriting. After the first test the average speed of the class was equal to, or higher than, the best average of the preceding semester. The same is true of the three poorest typists in the class. The three best typists actually gained during the vacation period, as every test taken after the summer vacation period gives a distinctly higher average than that of the best tests for the same pupils the preceding semester.

Neither did the summer vacation show any marked effect on the accuracy of typewriting. The average accuracy for the class, Chart II, is the same for the first test in the fall as for the last one in the spring. The scores on all of the remaining tests for this group are fairly constant. The average score of the three most accurate typists was almost the same the first fall test as on the last test in the spring and remained fairly constant thereafter. The three least accurate typists still showed signs of being irregular but not quite so much so as during the first semester.

Table I brings out the fact that more pupils wrote with a greater speed on the first test in the fall than they did on the last test in the spring. By the time the fifth test was given only one had failed to better his mark of the preceding semester. In accuracy, Table II, about the same number did not improve as improved.

The general conclusion then, from this study, seems to be that a summer vacation period had little or no deleterious effect in either speed or accuracy on these pupils who had had a half year's work in typewriting, and that pupils continue to build up their speed from that point at which they left off the preceding semester.

TEACHERS' GRADES AS CRITERIA OF ACHIEVEMENT OF BI-LINGUAL CHILDREN

FLOYD F. CALDWELL
State Teachers College, Chico, California
and MARY DAVIS MOWRY

The problem of pupil's grades is always one of serious concern to the conscientious teacher. This problem in all probability becomes even more involved when a teacher has under her care a number of pupils who do not speak English as their native language and are thereby handicapped in responding to routine class-room procedures.

This study is an attempt to determine (1) whether teachers' grades assigned to these bi-lingual children are as fair criteria of school achievement as grades assigned to the English-speaking group; (2) whether there is a tendency for teachers to grade this bi-lingual group consistently higher or lower than knowledge of subject matter would justify.

This investigation is delimited to the study of only one foreign language-speaking group, namely the Spanish-American. The children belonging to this group are native born and have parents who have lived under the American flag for three generations but still speak Spanish as their native tongue. The group referred to as Anglo-American is composed of native English-speaking children.

Methods and Procedures

Objective and essay tests were constructed that were designed to test a knowledge of English and history. The test items and questions were selected from the text-

books that were in use in the different grades and in each case the essay tests covered the same subject-matter as was covered by the corresponding objective test. Also, the essay tests carried the same number of points as the corresponding objective tests; that is, it was possible for the pupils to earn the same scores on the essay as could be earned on the corresponding objective test. Largely due to this fact the essay tests were more detailed and corresponded more closely to objective tests than is true of the ordinary run of class-room essay tests. This factor was conducive to more objectivity in grading the essay tests but, on the other hand, the differences which appeared in the size of scores earned by the two racial groups on the two types of tests may have been considerably reduced.

A number of forms of each type of test were constructed. These were then administered by the class-room teachers as a part of the regular class-room routine. The objective tests were administered first in each case so that if any practice effect occurred the essay test results are the ones affected. All of the tests were checked and graded twice, the essay type being graded by more than one grader in order to control the subjective factor as much as possible. Furthermore, the grading was done by disinterested parties. Precautions were taken so that the graders were kept in ignorance as far as possible as to the nationality of the child whose paper was being graded. The scores from each type of test were then pooled to form what would be the equivalent of results obtained on one long objective and one long essay test for each grade. Because of this fact the reliability of the measures should be materially increased. A total of 167 Anglo-Americans and 216 Spanish-Americans were tested in English while 153 Anglo-Americans and 196 Spanish-Americans were tested in the field of history. A total of 3000 short tests were administered.

Analysis of Data

Table I gives the correlation coefficients obtained when the scores earned by the two language groups on the objective and essay tests were correlated with those earned on the New Stanford Achievement Tests. It will

TABLE I

Correlation of Objective and Essay Test Scores and Those Earned on the Stanford Achievement Test

Grade	Kind of Test	Anglo-American			Spanish-American		
		r	PE	N	*r*	PE	N
ENGLISH							
Fourth	Objective	.598	.094	21	.631	.064	40
	Essay	.651	.084	21	.441	.085	40
Fifth	Objective	.758	.042	46	.779	.032	68
	Essay	.678	.034	46	.456	.065	68
Sixth	Objective	.843	.029	46	.801	.035	46
	Essay	.798	.036	46	.491	.075	46
Seventh	Objective	.775	.044	25	.936	.150	32
	Essay	.889	.028	25	.588	.078	32
Eighth	Objective	.785	.047	29	.673	.067	30
	Essay	.688	.066	29	.634	.072	30
HISTORY							
Fifth	Objective	.710	.048	48	.668	.042	77
	Essay	.758	.041	48	.382	.065	77
Sixth	Objective	.852	.028	44	.861	.023	57
	Essay	.817	.034	44	.436	.074	57
Seventh	Objective	.852	.033	31	.842	.035	32
	Essay	.746	.054	31	.495	.090	32
Eighth	Objective	.781	.048	30	.728	.058	30
	Essay	.883	.027	30	.385	.108	30

be noted that the coefficients obtained by correlating the objective test and the New Stanford Achievement Test scores correspond very closely for the two language groups. There is no consistent tendency for the coefficients of one group to be higher or lower than those of the other. However, examination of the coefficients obtained by correlating the essay test and the New

Stanford Achievement Test scores reveals marked differences. These coefficients for the Spanish-American group are consistently much lower than those obtained by the same group with the objective and New Stanford Achievement Test correlations. Further, these coefficients are the lowest of any obtained in the series. Due to the fact that the New Stanford Achievement Test is a highly reliable measure, these coefficients would indicate (1) that the objective type of test is a more reliable measure of general school achievement for Spanish-American children than is the essay test, (2) that essay tests are a more reliable measure of general school achievement for Anglo-American than for Spanish-American children.

Table II reveals that for the Anglo-Americans the correlation of scores earned on each of the two types of tests with teachers grades produces fairly high coefficients. The essay test coefficients, however, are in this case lower than those for the objective test. On the contrary, with the Spanish-Americans when corresponding scores are correlated, the essay test coefficients are the larger. If, as is indicated in Table I, the essay test scores are of lower reliability, there seems to be reason to conclude that the teachers grades for the bi-lingual group are not as reliable criteria of achievement as those grades assigned to the Anglo-American children.

A further analysis of the tables reveals another interesting condition. Whereas the coefficients obtained by correlating the objective test and New Stanford Achievement Test scores of the Spanish-Americans, are as high or higher than those obtained for the Anglo-Americans, the coefficients obtained by correlating the objective test scores and teachers grades for the Spanish-Americans are consistently lower than those for the Anglo-Americans. This seems to indicate that there is a constant bias or a definite "halo" effect in evidence and that the Spanish-American children are being graded

TABLE II

Correlations Between Test Scores and Teachers' Grades

Grade	Kind of Test	Anglo-American			Spanish-American		
		r	PE	N	*r*	PE	N
ENGLISH							
Third	Objective	.894	.022	38	.818	.036	37
	Essay	.576	.073	38	.894	.022	37
Fourth	Objective	.912	.025	21	.690	.056	40
	Essay	.783	.056	21	.805	.038	40
Fifth	Objective	.791	.062	46	.648	.047	68
	Essay	.875	.024	46	.889	.017	68
Sixth	Objective	.713	.049	46	.653	.057	46
	Essay	.699	.051	46	.723	.047	46
Seventh	Objective	.853	.036	25	.821	.039	32
	Essay	.762	.057	25	.819	.039	32
Eighth	Objective	.798	.045	29	.678	.068	30
	Essay	.764	.052	29	.743	.055	30
HISTORY							
Fifth	Objective	.750	.042	48	.638	.045	77
	Essay	.630	.061	48	.709	.038	77
Sixth	Objective	.843	.028	44	.584	.058	57
	Essay	.743	.046	44	.838	.062	57
Seventh	Objective	.877	.028	31	.707	.060	32
	Essay	.779	.074	31	.846	.034	32
Eighth	Objective	.813	.045	30	.632	.046	30
	Essay	.733	.057	30	.832	.038	30

consistently lower than their actual knowledge of subject matter justifies.

Although this study is limited in that it includes only one bi-lingual group, namely the Spanish-American, it seems reasonable to expect that somewhat the same condition may exist in mixed classes having other types of foreign language-speaking children. Perhaps further study of these mixed class groups could determine the nature of the biases, prejudices, emotional or sentimental influences, et cetera, that are in operation under these conditions and further, these factors might be studied in order to reveal the degree to which they operate to influence class-room methods and procedures.

A PERSONALITY STUDY OF A GROUP OF HIGH SCHOOL HONOR SOCIETY PUPILS

T. L. ENGLE

Isaac C. Elston Senior High School, Michigan City, Indiana

Probably the greatest academic honor which can be bestowed on high school pupils is election to membership in their local chapter of the National Honor Society. The present study is an attempt to analyze some of the personality "traits" of a group belonging to this society and to find to what extent these traits correlate with scholarship. The study has been divided into three parts:

(1) Intelligence as shown by an intelligence test and its relation to scholarship.

(2) Personality traits as shown by teacher ratings and their relation to scholarship.

(3) Personality traits as shown by a standardized personality inventory and their relation to scholarship.

The subjects of the study were twenty-four members of the Isaac C. Elston Senior High School, Michigan City, Indiana, chapter of the National Honor Society for the school year 1931-32.

By the rules of the society four criteria are used in selecting members: scholarship, character, leadership, and service. To be eligible for election a pupil must be in the upper 25th per cent of his class as shown by the regular school marks for his entire time in high school. From this selected group 15 per cent of the seniors and 5 per cent of the juniors may be elected to membership although it is the policy of the school not

to elect the full quota. In order to select the pupils who meet the other three requirements each teacher is given a list of eligible pupils and is asked to rate those known to him on character, leadership, service, and general desirability for membership. The results are tabulated and then with this help a faculty committee selects the pupils to be honored. Undoubtedly the selection of members on all points except scholarship is very subjective and open to criticism, but until recently there have been no objective methods for judging personality traits, and even at the present time such methods are in the experimental state and are not well known and accepted.

1. Intelligence as Shown by an Intelligence Test and Its Relation to Scholarship

The members of the group were given the *Otis Self-Administering Tests of Mental Ability, Higher Examination, Form D,* 30 minutes being allowed for the test. The scores gave IQ's ranging from 90 to 121 with a median of 111.5, an upper quartile of 116, and a lower quartile of 105. The IQ's were correlated with scholarship as shown by school marks, high IQ's and high scholarship being ranked together. A correlation of positive .182 with a probable error of .139 was found. Evidently intelligence as shown by an intelligence test was not an outstanding qualifying factor in scholarship so far as this highly selected group was concerned, although the group as a whole stood high in IQ.

2. Personality Traits as Shown by Teacher Ratings and Their Relation to Scholarship

In the Michigan City schools each pupil in the Junior and Senior High Schools is rated by at least three teachers each year on ten characteristics. For the group under study, ratings were available for the full six

years in the two schools. Of course the ratings are open to the usual criticisms of ratings, but they do represent the best efforts of the teachers to judge qualities other than scholarship. The ratings were weighted and these weighted scores were correlated with scholarship.

TRAIT	r	PE_r
Industry	.522	.105
Punctuality	.479	.111
Ambition	.387	.123
Dependability	.380	.123
Initiative	.134	.141
Common Sense	.116	.142
Courtesy	.013	.144
*Honesty	.000	.144
Self-Reliance	—.021	.144
Appearance	—.226	.137

* All pupils were given the highest possible rating by all teachers.

The negative correlation between appearance and scholarship, although slight, is of interest in that it reminds one of the comments sometimes heard on college campuses that the Phi Beta Kappas are the most poorly dressed people in school. Evidently for this highly selected group there is very little relationship between any characteristic and scholarship, although industry, punctuality, ambition, and dependability seem to be somewhat significant factors in scholastic achievement.

3. Personality Traits as Shown by a Standardized Personality Inventory and Their Relation to Scholarship

In order to secure a more objective basis for studying the personalities of the members of the Honor Society *The Personality Inventory* by Dr. Robert G. Bernreuter was given to them. The resulting scores were correlated with scholarship.

A. Neurotic Tendency and Introversion-Extroversion.

The B1-N scale of the Bernreuter Inventory measures neurotic tendency and the B3-I scale measures introversion-extroversion. Although the two scales can be scored separately Dr. Bernreuter suggests that in as much as the inter-correlations are very high only one scale need be scored. In the present study the B1-N scale was used. In making the correlations with scholarship high scores on neurotic tendency, that is, tending to be emotionally unstable, were ranked with high scholarship. This automatically made the most introverted scores ranked with high scholarship. A correlation of negative .064 with a probable error of .143 was found.

B. Self-Sufficiency.

The B2-S scale is a measure of self-sufficiency. Dr. Bernreuter, in his manual accompanying the Inventory, says of this scale, "Persons scoring high on this scale prefer to be alone, rarely ask for sympathy or encouragement, and tend to ignore the advice of others. Those scoring low dislike solitude and often seek advice and encouragement." In correlating self-sufficiency with scholarship high scores were ranked with high scholastic marks. A correlation of positive .180 with a probable error of .139 was found.

C. Dominance-Submission.

The B4-D scale of the Inventory is a measure of dominance-submission. In making the correlation with scholarship the scores showing the highest degrees of dominance were ranked with highest scholastic marks. A correlation of positive .119 with a probable error of .142 was found.

Thus it seems that for this highly selected group of Honor Society pupils there is no single factor which correlates significantly with scholarship.

VARIATIONS IN JUDGMENT

FRANCES V. MARKEY
Teachers College, Columbia University

This study deals with the relation between consistency and accuracy of judgments and the resemblance between judgments in one situation as compared with another. The study is based upon students' judgments of the age and intelligence of photographed individuals whose actual age and intelligence (as measured by mental tests) were known to the investigator.

A review will indicate the scope of methods used in investigations of similar problems. Moriwaki (6) found that judgments from photographs are only slightly better than chance in accuracy. Gaskill, Fenton and Porter (3) secured a group correlation coefficient of .70 in estimates of intelligence from photographs of boys, while the median r (of the individual coefficients of the judges) was only .425; Knight (4) likewise found that group judgments were more accurate. Anderson's (1) subjects separated photographs into four groups according to mentality; the average correct placement was low.

Laird and Remmers (5) had subjects sort uniformly sized pictures of college students photographed in academic costume; a chance correlation with intelligence test scores was obtained. Seven judges showed no consistency in ability when retested with similar material; a character "expert," however, did no better than the average judge. Cogan (2) compared the ratings of associates and non-associates for snobbishness, vulgarity, intelli-

The author wishes to acknowledge the assistance of Professors G. S. Gates and H. L. Hollingworth of Barnard College in formulating the problem and obtaining the data of this study.

gence and other traits. Non-associates' estimates, based on photographs, correlated highly with those of associates, formed through acquaintance, in the case of all traits except intelligence. Since objective criteria of the traits were not available, Cogan did not accept the identity of the judgments as proof of their validity. Pintner (7) found that psychologists were slightly better than physicians, business men and teachers in judging intelligence from photographs. One subject, when retested after six months, showed less accuracy than on the first trial.

The materials used in the present investigation were the twelve photographs prepared by Pintner (7). On two occasions, separated by an interval of two weeks, the subjects, members of elementary (Group A) and advanced (Group B) psychology courses at Barnard College, were required to rank the photographs according to age and intelligence. The concept of intelligence quotient was explained so that judgments would be of relative rather than of absolute capacity. Number one was assigned to the most intelligent individual and number twelve to the least, number one to the oldest and number twelve to the youngest, and the others were ranked accordingly. Subjects were instructed to choose the most intelligent (or the oldest) and to place the picture face down, then to choose the next most intelligent, and so on, and to record their judgments on standard forms.

Group A was allowed 40 minutes and Group B 20 on both occasions. On the second trial both groups used only half the allotted time in most cases. Instructions were given at both sittings. The purpose of the experiment was not revealed until the end of the second trial, and the subjects had not been told that a second trial would be given; they were informed merely that scores from the first sitting would be given at a later time. Records were discarded in the case of subjects to whom the photographs were familiar. In some respects the

materials did not provide for maximum accuracy, since the age and intelligence differences of the individuals photographed were not equal in amount on a quantitative scale, and since both sexes were pictured; further, the clothing was outmoded and the poses of the subjects varied considerably. One or two judges inquired whether all were equally intelligent; it was therefore explained to all that no "trick" was involved. The experimental errors were substantially the same for all subjects, hence the group results may be considered valid.

The table below shows the mean rank difference correlation coefficients (with SD's) between the students' judgments of intelligence and the actual intelligence ranks of the photographed individuals for the first trial under I_1, for the second under I_2; the corresponding figures for age judgments are under A_1 and A_2. Under V_1 are given the mean rank difference coefficients between the students rankings of intelligence on the first trial and their corresponding rankings on the second trial, and under V_a the parallel results for age.

TABLE I

Average Rank Difference r's, with Sigmas

	I_1	I_2	A_1	A_2	V_1	V_a
		Group A	N = 47			
Average	20.1	20.6	71.1	72.1	85.9	93.1
S.D.	29.9	34.7	10.6	11.7	12.6	8.9
		Group B	N = 36			
Average	18.1	24.2	74.8	73.0	80.8	90.5
S.D.	30.1	26.7	10.9	11.9	23.1	10.7

In Table II the product moment correlations are shown between scores (as represented by the correlation between each subject's judgments of age or intelligence and the actual age or intelligence of the photographed individuals) on the first and second trials in judgments of intelligence (I_1I_2) and age (A_1A_2); between scores

in judgments of age and intelligence on the first trial (I_1A_1), and the second trial (I_2A_2). The table also shows the correlations between consistency of judgments and accuracy of judgments of intelligence on the first trial (I_1V_1) and on the second trial (I_2V_a); the corresponding figures for age are shown under A_1V_a and A_2V_a; V_1V_a is the product moment of the consistency coefficients for age and intelligence. The P.E.'s of the coefficients are given; the differences between the coefficients obtained in the two groups with probable errors are also indicated.

TABLE II

Product Moment r's, P.E.'s, and Differences with P.E.'s

	I_1I_a	A_1A_2	I_1A_1	I_2A_2	I_1V_1	I_2V_1	A_1V_a	A_2V_a	V_1V_a
				Group A	N = 47				
Coefficient	.849	.589	−.095	.072	.128	−.025	.229	.304	.288
P.E.	.027	.064	.097	.097	.096	.098	.093	.089	.089
				Group B	N = 36				
Coefficient	.517	.590	.014	.131	.126	−.446	.066	.574	.020
P.E.	.082	.063	.129	.110	.111	.090	.112	.065	.112
Difference	.332	.001	.109	.059	.002	.421	.163	.270	.268
P.E. Diff.	.086	.089	.163	.147	.146	.133	.145	.110	.143

Consistency in judgment in one situation does not appear to be related to consistency in another, when the correlation between consistency in judgments of age and judgments of intelligence made at disparate intervals is considered; although the difference between groups A and B is not reliable it is noteworthy that the group which had the longer time interval in which to make the judgments obtained a slightly higher coefficient of consistency .288, as compared with .020. Memory perhaps tended to favor those given the longer period in which to make judgments.

The relation of consistency to accuracy in judgment may be obtained from the coefficients for the first and second judgments of age and intelligence with their in-

dices of consistency. Judgments of intelligence correlate on the whole only slightly with their measure of consistency, the judgments on the second trial being related inversely to a somewhat marked degree in Group B. In the case of age, a positive but insignificant relation between consistency and accuracy is revealed for both trials. Again, the average of the coefficients of the group with the shorter time interval is somewhat lower. Considering the coefficients as a whole, the results do not show a high degree of relationship between consistency and accuracy in judgment; similarly, accuracy in one judgment situation does not appear to be related to accuracy in another, as shown in I_1A_1 and I_2A_2 coefficients. A composite score would not alter the results inasmuch as the coefficients are so small.

Another aspect of consistency of judgment as related to accuracy of judgment is given in the first two columns of Table II. The difference between the groups is significant but not reliably so in the case of judgments of intelligence, and is neither significant nor reliable in the case of age. The effect of memory in the longer time interval may be operative here; the time interval may affect the consistency of judgment but it does not affect the accuracy, since the groups do not differ greatly in mean score for the two trials (Table I). This might indicate that in the case of judgments of intelligence, which in general are more difficult than judgments of age, a longer time interval possibly aids the consistency of the judge by facilitating the retention of cues but does not improve the choice of cues on which judgment is based. This is indicated also by the figures for comparative variability of the two groups in consistency of judgments of intelligence, the slow group having a coefficient of variability of 14.7 and the fast group of 28.6.

Correlations between scores on first and second trial are more nearly identical for both groups in the judgments of age. Comparative variabilities are not signifi-

cant in the case of the consistency of these judgments. It might be concluded that the criteria of age in photographs of individuals, though not valid in the light of objective data, are well enough formulated by the average person to be relatively free from the influence of time in making judgments. In the more difficult situation, i. e., in judging intelligence, time perhaps aids consistency of judgment, since the cues are possibly not so clearly formulated.

The validity of judgments of age and intelligence, it appears, is far from adequate when made on the basis of photographs, but more adequate in the case of age. This confirms previous evidence of the uselessness of photographs in estimating intellectual status.

In summary then, general consistency in judgment is not apparent in the somewhat similar situations of estimating age and intelligence from photographs. Accuracy in judging does not correlate highly with consistency, nor is accuracy in one situation related to accuracy in the other. Consistency in successive judgments suggests objectivity when of a high degree, and subjectivity when low; group variability and the conditions of the judgment (time) must be considered in interpreting consistency. Judgments of age and intelligence from photographs correlate poorly with objective criteria, less poorly in the case of age. A longer period of exposure tends to increase consistency in successive judgments, and to decrease group variability in consistency to some extent, but does not improve the accuracy of judgments.

1. Anderson, L. P. "Estimating Intelligence by means of Printed Photographs." *Jour. of Applied Psych.*, 1921, 5, 152-55.

2. Cogan, L. G. "Judgments of Traits of Character on the Basis of Photographs." Columbia University, Master's Essay, 1916.

3. Gaskill, P. C., Fenton, N. and Porter, J. P. "Judging the Intelligence of Boys from Their Photographs." *Jour. of Applied Psych.*, 1927, 11, 394-403.

4. Knight, H. C. "A Comparison of the Reliability of Group and Individual Judgments." Columbia University, Master's Essay, 1921.

5. Laird, D. A. and Remmers, H. A. "A Study of Estimates of Intelligence from Photographs." *Jour. of Experimental Psych.*, 1924, 7, 429-46.

6. Moriwaki, E. "Note on the Comparative Validities of Judgments of Intelligence Based on Photographs and on Interviews." *Jour, of Applied Psych.*, 1929, 13, 630-36.

7. Pintner, R. "Intelligence as Estimated from Photographs." *Psychological Review*, 1918, 25, 286-296.

AN HISTORICAL NOTE ON THE APPLICATION OF PSYCHOLOGY

C. M. LOUTTIT
Indiana University

It can hardly be denied that psychological principles have been applied since the dawn of man's intercourse with his fellows, yet the history of intentional application of such principles to the affairs of the everyday world is probably very short. Except in the field of education the history would cover perhaps but very little more than the first quarter of the twentieth century. The late eighteenth century has, however, an interesting contribution best described by its title page.

> A / SPECIMEN / OF / A COMMENTARY / ON / *SHAKSPEARE.* / CONTAINING / I. / NOTES ON AS YOU LIKE IT. / II. / AN ATTEMPT / TO / EXPLAIN AND ILLUSTRATE / VARIOUS PASSAGES, / ON A NEW PRINCIPLE OF CRITICISM, / DERIVED FROM / MR. LOCKE'S DOCTRINE / OF / *THE ASSOCIATION OF IDEAS.* / (line) / London: / PRINTED FOR T. CADELL, IN THE STRAND. / 1794.

This work is the product of Walter Whiter and was written and published while he was a Fellow of Clare College, Cambridge. He was later a clergyman and published several works of some importance in philology. As stated on the title page the author has taken a definite psychological principle and applied it to the field of literary criticism. This sort of application is not unknown today if one considers the work of the psychoanalysts and others.

The author's argument is that "as it is the business of the critic to discover and establish the original language of the author, and to reject what is sometimes called the *improved* text of an *ingenious* commentator, we shall instantly perceive, that from this principle may probably be derived a very important canon for the confirmation of disputed readings, which have perhaps been too hastily condemned, as quaint, remote, or unin-

telligible." Thus the use of certain language or vocabulary may depend upon stimulating situations "peculiar to the country, the age, and the situation of the writer," as for example the masques and pageants of the time of Shakspeare.

The principle is applied *in extenso* to explain many passages in Shakspeare's plays and poems that earlier critics had either set aside as a play on words, or were unable to account for. Thus the source of the imagery of clouded heavens, smoke from Hell, the cave of death, and so on, Whiter finds in the stage settings and properties, and in the technical vocabulary of the theater of the day. The curtain loft being known as "Heaven", the lower stage as "Hell", the usual setting for tragedy being black draperies, and the frequent use of a cave sending forth smoke as the mouth of Hell; all of which the author finds from other sources were characteristic of the Elizabethan stage.

Frequent allusions to classical times need not indicate that the author toiled "in the study or perusal of books." Rather we should, on the basis of the association principle, expect them to occur as the times were replete with masques and pageants usually dealing with classical themes. The artist saw and heard all of these and quite unconsciously they have influenced his subsequent work.

To the modern reader some of the examples may seem remote and the explanations not always clear, yet in several instances the author has apparently made out a good case for his resolution of disputed points. In one instance—the only one in which Shakspeare is not involved—the results do not conform to subsequent conclusions. The famous discussion regarding the poems of Rowley, which Chatterton purported to have transcribed, is settled according to the association principle, in favor of Chatterton's contention. That this youth with his poor education could have written in ancient phraseology evident in portions of the poem, Whiter thinks unlikely; he then further demonstrates that just that phraseology should be expected from a medieval writer. And Rowley was supposed to be a medieval monk.

This is not the place, however, to discuss the merits of Whiter's criticism or method. It does seem important as well as interesting to call attention to this work as probably the earliest intentional application of a psychological principle in another field of endeavor.

NEWS AND NOTES

The Psychological Institute of the University of Vienna will hold its third annual summer school in psychology for American students from July 9 until August 9, 1934. The courses, which are taught in English, include the following: Language and Personality (Karl Buehler), Childhood and Adolescence (Charlotte Buehler), Biographical Methods (Charlotte Buehler), Viennese Tests for Children, Experimental Psychology, Business and Social Psychology. The University of Kentucky is again supervising these courses and will grant six semester-hours university credit for the work done in Vienna. For further details address the Educational Director, Dr. Henry Beaumont, University of Kentucky, Lexington, Kentucky.

A symposium on "The Analysis of the Individual," prepared by twelve of the outstanding psychologists of the United States and Great Britain under the editorship of Donald G. Paterson, appears in the April issue of "Occupations, the Vocational Guidance Magazine." Those interested may secure single copies for fifty cents, postpaid, from National Occupational Conference, 522 Fifth Avenue, New York.

Two years ago the National Council of Teachers of English started an experiment to determine whether the motion picture tastes of high school pupils could be improved through the medium of the English class. The experiment was so successful that at the annual convention of the National Council in Detroit last December, the thousand teachers present voted the adoption of the report of William Lewin, chairman of the Photoplay Appreciation Committee, recommending that photoplay instruction be introduced in the schools and that courses of methods in teaching photoplay appreciation be given in schools of education. Already curriculum commissions in New York, Virginia, Maryland and California have recommended the inclusion of such courses in junior and senior high schools.

BOOK REVIEWS

MOTION PICTURES AND YOUTH: A REVIEW[1]

C. M. LOUTTIT, *Indiana University*

Motion pictures have, during their short history of scarcely more than thirty years, become established as one of the most important influences of modern society. With the possible exception of the newspaper, they probably touch the lives of more people than any other of the major institutions of our times. It is probable that, within the limits of their content, they influence more people than even the newspapers because the latter is available only to literates, while pictures may appeal to all except the relatively small number of blind. However, no one can say just how much social influence the motion picture has had, although the volume of writings on the subject will total to hundreds of titles. Opinion, bias, preconceived notions, wish-fulfillment, and all the gamut of error-producing agents, have characterized many, if not most, of these.

Now we have an extensive study of the influences exerted by the movies on children and youth which is remarkably free from bias. "This series of twelve studies of the influence of motion pictures upon children and youth has been made by the Committee on Educational Research of the Payne Fund at the request of the National Committee for the Study of Social Values in Motion Pictures, now the Motion Picture Research Council, 366 Madison Avenue, New York City. The studies were designed to secure authoritative and impersonal data which would make possible a more complete evaluation of motion pictures and their social potentialities." Surely the data are "authoritative and impersonal," they contribute much to the evaluation of motion picture influences, and they afford an excellent body of material on child behavior.

Eighteen authors—five educationists, four sociologists, and nine psychologists (the classification is the present writer's)—have contributed to this study. Each has approached his or their specific problem with different techniques; has used different subjects; large and small numbers of subjects of different ages and

[1] Publications of the Indiana University Psychological Clinics, Ser. II, No. 10.

sexes, and has presented the results in an independent book. This is perhaps the best way the study could be conducted. But a careful summary integrating the various findings is absolutely necessary. This has been in part supplied by Charters (1), who as chairman of the committee, summarized the reports. One could wish for a little more critical summary by someone who is not bound by a chairmanship to accept all as good. His three "inevitable conclusions" are vague enough and innocuous enough to cause no violent opposition. These conclusions are: (1) "the motion picture, as such, is a potent medium of education;" (2) "for children the content of current pictures is not good," and (3) "the motion picture situation is very complicated. It is one among many influences which mold the experience of children." If these conclusions truly represented the contribution of the research of eighteen people, then the Payne Fund has invested in proof of the obvious.

It does not seem to the present writer that these three conclusions even remotely tell the story of the findings of this research program. While it is patently impossible to survey each work in detail, perhaps even brief accounts may suggest that the real conclusions are far more serviceable than those given as "inevitable" by the chairman. For convenience in discussion, the studies may be divided into: (1) three which may be considered more or less as laboratory experiments; (2) two studies of attitudes; (3) two studies of conduct using a case study technique, and (4) two of the books which are not directly concerned with the influence of moving pictures on children.[2]

Renshaw, Miller and Marquis (7) have made an exhaustive study of children's sleep and the factors which affect it. The experiments were conducted on 92 boys and 79 girls from five to nineteen years of age, all of whom were living at the Ohio Bureau of Juvenile Research. Thus the activities of the group could be carefully controlled. Measurements were made of bodily motility during sleep by adequate recording devices attached to experimental beds in which these children slept.

Careful study of the normal or usual sleep motility showed that for a given child the hourly motility distribution is in a high degree characteristic. The pattern determined over a period of 10 to 15 nights can be used as a basis for comparing the influence of unusual activity during the day on the child's sleep. Three sorts of unusual activity were investigated for their influence on the sleep motility curve: coffee, loss of sleep, and moving pictures. Expressed in terms of group averages these results

[2] See note at beginning of list of references.

indicate that coffee drinking tends to increase sleep motility, experimental insomnia tends to decrease it, while moving pictures do both, although increased motility is more usual, especially in older children.

Results based on group averages, as is emphasized again and again by the authors, may be very misleading. Thus in relation to movie influences they say, "This makes *group* averaging a poor representation of the true situation. We are convinced. * * *, that movie influence is a highly individual affair, and that each child presents a problem such that we can only study *how he differs from himself* when we look at him under the normal and the experimentally varied conditions." (7, p. 135.) Again in relation to coffee drinking: "Individual differences in susceptibility to the influence of coffee are so great that it is impossible to generalize regarding the extent or direction of its effects on children of the same sex or of the same age." (7, p. 207.)

In contrasting the influence of these two factors, the following quotation is significant: "Parents who would strongly protest against their young children ingesting from four to six grains of caffien [equivalent to two to three cups of coffee] between the hours 6:00 and 9:00 P. M., nevertheless permit attendance at motion pictures whose effects on sleep motility may be as great or greater than that of coffee and possibly more lasting in influence." (7, p. 208.)

A subsidiary study by Miller (16) reported in this volume shows that the critical flicker limen for children of either sex is sufficiently low that the rate of motion picture projection should have no effect in causing eye-strain.

Dysinger and Ruckmick (5) using the psychogalvanic technique studied the "emotional" responses of children and adults when observing motion pictures, both in the laboratory and in the theater. Responses to scenes of conflict and danger were most intense in children under 12 years of age and decreased with advancing age; responses to scenes of love and sex were of least intensity at this low age, but most intense in adolescence. In this study individual differences also loom large and the authors are led to say, "Yet it is clear that in the last analysis attendance at the motion-picture theater is a matter of individual mental lives and must be regulated, or at least judged, according to the individual psychophysiological organism. Generalizations may be made on the basis of average results, but the question resolves itself into a personal equation: the person who attends the theatre in terms of his peculiar mental and physical constitution." (5, p. 115.)

The authors' last paragraph hints at the possible practical significance of the experimental findings: "Our records are clear on this point: profound mental and physiological effects of an emotional order are produced. The stimulus is inherently strong and undiluted by post-adolescent critical attitudes and accumulated and modifying experiences. Unnatural sophistication and premature bodily stimulation will then result. The emotions are widely known in the psychological literature for their extensive and intensive grip on mind and body. Let those moralize and standardize who will: our fundamental facts speak loudly for themselves." (5, p. 119.) Blumer (8) cites nearly one hundred biographical accounts of emotional possession that vividly corroborate this suggestive statement.

The third of the experimental studies is concerned with the retention of material seen in moving pictures. Holaday and Stoddard (2) tested the retention of ideas over various periods of time in groups of children from the second to the tenth grade and in adults. The total number of subjects was over 3,000 and the content of some seventeen pictures were tested.

It was found, surprisingly enough, that young children in the second and third grades have about 60 per cent of the retention of a group of superior adults. Equally surprising was the retention after one month, "almost as high as in the original testing immediately after the picture," and that retention scores after three months were between 80 and 90 per cent of the immediate recall. And this in spite of the fact that the picture was shown only once under unfavorable conditions for learning and no previous warning was given that recall would be required. Retention was facilitated by the emotional effect of the scene and also by reactional biography of the individual. The effect of environmental background and reactional biography is shown in another connection by Blumer and Hauser (11) as is discussed below.

The educational value of moving pictures is indicated by the increase in general information after seeing pictures in which information tested was correctly shown, but when information was incorrectly shown the general information scores decreased even in the adult group. "The content of a picture is accepted as authentic by a large percentage of the audience unless the errors contained are glaring." (2, p. 78.)

Ideas secured from movies are influenced by the age of observers. While children do understand pictures the younger ones do not comprehend the plot and they are apt to remember inconsequential details. A similar circumstance was found by Renshaw, Miller and Marquis (7) who say: "The younger children

recalled incidents during the interviews which were largely drawn from the comedies. As a rule the younger children did not get a continuity of story from the feature picture." Similarly Dysinger and Ruckmick (5) found wide differences in perception, especially between the below-12-year-group and adults. They feel that "errors of perception at this age must radically affect the understanding of the story." (5, p. 111.)

The problem of measurement of social attitudes and the influence of various factors in changing such attitudes has been attacked with eminent success by Professor Thurstone who published the method some years ago (17). In the present work Peterson and Thurstone (3) republish certain reports on the effect of movies on social attitudes together with much new material. From their results there can be no question that movies do influence the attitudes of adolescents toward other nationalities and races (Germans, Chinese, and Negroes), gambling, war, and the punishment of criminals, all of which were investigated. The change in attitude may be more or less favorable depending upon the point of view taken in the picture. Persistence of the change in attitude was found even up to nineteen months. Some pictures when shown singly were found not to influence attitudes, but two or more such pictures with similar points of view acted in summation so that a change in attitude was found after viewing them.

Another study of attitudes together with conduct of children as influenced by the movies is reported by Shuttleworth and May (4) who used the elaborate questionnaire and statistical analysis technique earlier employed in the Character Education Inquiry (15). In essence the plan was to compare matched groups of "movie" children and "non-movie" children—*i. e.*, those who attended movies two or more times a week, and those who attended movies once a month or less—in a wide variety of measures of conduct, attitudes, and school adjustment. "Of the total number of comparisons made between the movie and non-movie groups including conduct tests, school grades and deportment, attitude tests and questionnaires, significant differences favoring the non-movie group appeared in about 8 per cent of the comparisons, favoring the movie groups in about 2 per cent, and favoring neither (*i. e.*, no difference) in about 90 per cent." (4, p. 84.)

Such a small percentage of comparisons between the two groups showing significant differences would suggest that perhaps movie influence is not operative in many areas of children's behavior. The authors suggest four hypotheses to explain the differences found, but feel that only two of them are supported by their data.

It is hardly necessary to recite their hypotheses here, but the last paragraph of their concluding chapter is significant. "That the movies exert an influence, there can be no doubt. But it is our opinion that this influence is specific for a given child and a given movie. The same picture may influence different children in distinctly opposite directions. Thus in a general survey such as we have made, the net effect appears small. We are also convinced that among the most frequent attendants the movies are drawing children who are in some way maladjusted and whose difficulties are relieved only in the most temporary manner and are, in fact, much aggravated. In other words, the movies tend to fix and further establish the behavior patterns and types of attitudes which already exist among those who attend most frequently." (4, p. 92.)

More direct influence of the movies on conduct was investigated by Blumer (8) and by Blumer and Hauser (11). In both studies the method used was essentially a careful study of autobiographies supplemented with questionnaires and interviews.

In the first of these studies, autobiographies emphasizing the movie experiences of the writers were secured from about 1800 high-school and college students. These were supplemented with a small number of interviews and reports of conversations. A questionnaire study of the influence of movies on conduct was made of 1200 fifth and sixth grade children. These data have not been treated statistically except occasional mention of percentage of responses. Instead of a statistical analysis, which would be rather unsatisfactory, about 100 selected quotations are given from the autobiographies. These cases illustrate the influence on childhood play, on adolescent imitation, on day-dreaming, on emotional possession in the way of fear, sorrow, love and passion, and excitement, on emotional detachment, and on schemes of life. Without doubt the cases cited do illustrate the positive influence of moving pictures. That the influences were not limited to those individuals whose histories are quoted is suggested by the percentage figures given. For most of the sorts of influence mentioned there were positive responses from large percentages of the persons submitting autobiographies, usually these positive percentages were between 50 and 70. However, it must be remembered that these influences were immediate, and it is questionable how long the effect may last. The present data cannot answer such a question except in relation to childhood play. Here the high school and college students remember that their play was influenced by movies but they do not feel that their subsequent behavior has been affected. That the immediate influence of

moving pictures is probably affected by the observer's reactional biography is indicated in the following quotation: "By skillful and dramatic presentation which grasps the attention of the observer and plays upon his impulses, motion pictures seem able to lessen reflective judgment and ordinary self-control, to stir impulses and facilitate their expression. This is most pronounced, of course, in the case of spectators who are already disposed to act along certain lines. The nervous child, the romantically inclined young girl, the young man "hair-triggered" on sex, the excitable boy, the boy who wants to be daring and "tough," persons with a bad or a sad conscience—all these, as the instances suggest, in witnessing pictures which play upon their respective tendencies are particularly likely to yield to incited impulse." (8, p. 126.) Occasionally, at least, moving pictures may have serious and persistent effect as is illustrated by the recent study of Brüel (14).

The autobiographical technique was also used by Blumer and Hauser (11) in studying the relation between movies and delinquency. Their subjects were juvenile and adult offenders and unselected groups of children from regions of high and low delinquency rates. While it is not possible to cite all of the figures given (although they are few, the argument being supported largely by quotation from autobiographies) one example is instructive. Seventeen per cent of behavior problem boys from a special school, nine per cent of boys in areas of high and medium delinquency rate, and only three per cent of boys from a low delinquency area reported positively to a question whether or not movies had led them "to do something wrong." Similar relative percentages were found concerning indirect influences of the movies. These would suggest again that the reactional biography of the child is probably of great importance in determining the influence of moving pictures. In the total group studied 10 per cent of males and 25 per cent of females attributed importance to movies as a factor in their careers. The authors feel these figures are "essentially conservative," even so, such small percentages suggest that probably motion pictures exert a small, though not negligible, influence on a delinquency in general.

A subsidiary study reported in Chapter IX on the use of movies in correctional and penal institutions indicates that they "may have appreciable recreational, disciplinary and reformatory value."

The remaining two volumes available at present in this series are not concerned with the influence of motion pictures on children. Peters (6) in the first of these presents a technique for measuring the mores for different types of social behavior, and then

rating motion pictures as coinciding or deviating from these. As this has a very indirect bearing on the problem of the influence of movies on children, we need do no more than mention it.

The last volume of the series by Dale (13) is a text book on motion-picture appreciation. While written for high-school students it is equally valuable for anyone who is unsophisticated in regard to motion picture production. Produced at an extremely small cost and sold widely—yes, even in the five and ten cent stores—this single book would have an influence greater than all the movie magazines put together. For the general public it is of greater value than all the other twelve volumes.

Having briefly noticed the books of this series, perhaps we can now suggest what may be more pertinent conclusions than those quoted earlier. One point recurs in report after report. That is the very important part played by the individual. Hardly a single author is entirely willing to stand by the conclusions based on average or group results. With no intention of distorting the researches, quotations have been given from several of the books emphasizing this point.

Moving pictures and coffee both affect sleep motility, but how they do so depends upon the individual. Moving pictures do effect emotional arousal, but the sort and degree depend upon the individual. Delinquency may in part be blamed to moving pictures, but not without considering the individual child. Moving pictures do influence the play of children and certain social responses of adolescents, but how and in what degree depend upon the individual. And so it could be said of every area investigated.

Rather than drawing conclusions as to the influence of moving pictures on children in general, we must emphasize the part they play in relation to the individual child. Interpreting the data from this point of view we may propose these conclusions:

(1) The influence of moving pictures in general, or of a specific picture on an individual child is determined in a great measure by the child's reactional biography. His physical environment, his early training, his companions, his home and family relationships, the models he has available for emulation, in short, the sum total of his experiences—interactions with such stimulus situations constitute the reactional biography—these are the things that determine in a large measure the influence of the movie on the child.

(2) Motion pictures cannot be commended nor condemned because of the influence they exert on *children*, but only because of their influence on this or that child. We cannot argue for reform of moving pictures because of their baneful effects on children,

or on adults for that matter. There are far better reasons to desire a change. But the data do argue for better guidance of the child. We do not frown upon automobiles because they frequently maim or kill a child. Rather, we train the child to use a sort of elementary judgment when crossing streets. Unfortunately, when it comes to matters of greater social import, as the movies, or books, or other interest, we leave the development of judgment to kind Fates, hoping for the best.

It is not possible to close this review without a word regarding certain editorial matters. Several of these reports give samples of questionnaires used (2, 3, 4, 5, 7, 11) and all (except 8 and 11) have adequate indexes. Holaday and Stoddard (2) have a bibliography of 36 titles and Renshal *et al.*, (7) have a classified bibliography of 255 titles. Charts, folding plates, and other similar aids to the presentation of technical material have been freely used. But consider this fact: the 13 books have been bound in nine volumes. Only five of the reports are separate volumes, the other eight being bound in pairs. Only one of these pairs has any justification for being bound together. Consider the quandary the library classifier is in when he tries to assign a classification to a book containing "The Emotional Responses of Children to the Motion Picture Situation," and "Motion Pictures and Standards of Morality!"

List of References

Part I

Note: At the time of writing this review, numbers 9, 10, and 12, were not yet published, or at least had not been received by the reviewer. In the bound volumes, numbers 1-2, 3-4, 5-6, and 9-10, are combined. The publisher in each case is the Macmillan Company of New York and the date is 1933.

(1) Charters, W. W. Motion pictures and youth. A summary. vii + 66 p.

(2) Holaday, Perry W. and Stoddard, George D. Getting ideas from the movies. viii + 102 p.

(3) Peterson, Ruth C. and Thurstone, L. L. Motion pictures and the social attitudes of children. xviii + 75 p.

(4) Shuttleworth, Frank K. and May, Mark A. The social conduct and attitudes of movie fans. v + 142 p.

(5) Dysinger, Wendell S. and Ruckmick, Christian A. The emotional responses of children to the motion picture situation. xiii + 122 p.

(6) Peters, Charles C. Motion pictures and standards of morality. v + 285 p.

(7) Renshaw, Samuel; Miller, Vernon L., and Marquis, Dorothy P. Children's sleep. xviii + 242 p.

(8) Blumer, Herbert. Movies and conduct. xiv + 257 p.

(9) Dale, Edgar. The content of motion pictures. (Not yet published.)

(10) Dale, Edgar. Children's attendance at motion pictures. (Not yet published.)

(11) Blumer, Herbert and Hauser, Phillip M. Movies, delinquency, and crime. xiii + 233 p.

(12) Cressey, Paul G., and Thrasher, Frederick M. Boys, movies, and city streets. (Not yet published.)

(13) Dale, Edgar. How to appreciate motion pictures. xi + 243 p.

Part II

(14) Brüel, C. Film als psycho-traumatisches Kindheitserlebnis (Psychotherapie). Acta psychiat, et neur. 1933, 8:445-454.

(15) Hartshorne, Hugh, and May, Mark A. Studies in the organization of character. New York: Macmillan, 1930.

(16) Miller, Vernon L. The critical frequency limen for visual flicker in children. In (7), Chapter ix, pp. 209-221.

(17) Thurstone, L. L. Attitudes can be measured. Amer. J. Sociol. 1928, 33:529-554.

NEW BOOKS AND PAMPHLETS RECEIVED

All books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

The Bundle of Life. JANE REVERE BURKE. E. P. Dutton & Company, New York, 1934. $1.50. 178 pp.

Cornell-Coxe Performance Ability Scale—Examination Manual. World Book Company, Yonkers, N. Y., 1934. $1.50. 88 pp.

Designed for Reading: An Anthology. Editors of The Saturday Review: HENRY S. CANBY, AMY LOVEMAN, WILLIAM ROSS BENET, CHRISTOPHER MORLEY, MAY LAMBERTON BECKER. Macmillan Company, New York, 1934. 614 pp.

Developmental Psychology. FLORENCE L. GOODENOUGH. D. Appleton & Company, New York, 1934. $3.00. 619 pp.

The Fundamentals of Psychology. Revised Edition. W. B. PILLSBURY. Macmillan Company, New York, 1934. $2.75. 663 pp.

Legal and Regulatory Provisions Affecting Secondary Education. WARD W. KEESECKER. Monograph 9, Bulletin 17, 1932, Office of Education, United States Department of the Interior, Washington, D. C. 114 pp.

Modern Man in Search of a Soul. C. G. JUNG. Harcourt, Brace & Company, New York, 1934. $3.00. 282 pp.

Der Primat des Geistes. VON ADOLPHE FERRIERE. Verlag Von Julius Beltz. Langensalza, Germany. 260 pp.

The Program of Studies. A. K. LOOMIS, EDWIN S. LIDE, B. LAMAR JOHNSON. Bulletin 17, 1932, Office of Education, United States Department of the Interior, Washington, D. C., 1933. 340 pp.

Secondary Education for Negroes. AMBROSE CALIVER. Bulletin 17, 1932, Monograph 17, Office of Education, U. S. Department of the Interior, Washington, D. C., 1933. 121 pp.

Short Stories of Today. RAYMOND W. PENCE. Macmillan Company, 1934. 512 pp.

A Social Basis of Education. HAROLD SAXE TUTTLE. Thomas Y. Crowell, New York, 1934. $3.00. 589 pp.

Social Psychology. ABRAHAM MYERSON. Prentice-Hall, Inc., New York, 1934. 640 pp.

Directory of American Psychological Periodicals

(Continued)

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Pedagogical Seminary and Journal of Genetic Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages, approximately. Edited by Walter V. Bingham. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Philosophy—New York: 515 West 116th Street. Subscription $4.00. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush, and H. W. Schneider.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Edited by Robert Saudek (London) and Ernest Seeman (Durham, N. C.). Founded 1932

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monograph—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300-400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $5. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Indianapolis: The C. E. Pauley & Co., Inc. Subscription $5.00. 600 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1917.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

(Continued on Third Cover)